1978

An exploratory study of teacher evaluation practices in selected Iowa schools

William Karl Tomhave
Iowa State University

An exploratory study of teacher evaluation practices in selected Iowa schools

by

William Karl Tomhave

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Professional Studies
Major: Education (Research and Evaluation)

Approved:

Signature was redacted for privacy.
In Charge of Major Work

Signature was redacted for privacy.
For the Major Department

Signature was redacted for privacy.
For the Graduate College

Iowa State University
Ames, Iowa
1978

TABLE OF CONTENTS

CHAPTER I. INTRODUCTION AND STATEMENT OF PROBLEM

Introduction

Teacher evaluation has become a frequent topic of debate within the public schools during the past few years, primarily in response to two forces: the rise of collective bargaining and a parallel rise in educational accountability due to rising taxes. Teacher evaluation itself is not new (Nutt 1920, Davis 1964), nor is the controversy regarding evaluation practices. To complicate matters, several states have mandated teacher evaluation apparently disregarding the existing problems (Redfern 1973, Sarthory 1973, Popham 1975b, Saif 1974). In 1976 the State of Iowa continued in this vein by enacting its own evaluation law. Section two hundred seventy-nine point thirteen (279.13) of the Iowa Code "CONTRACTS WITH TEACHERS--AUTOMATIC CONTINUATION," section three reads: "The board shall establish evaluative criteria and shall implement evaluation procedures. If an exclusive bargaining representative has been certified, the board shall negotiate in good faith with respect to evaluation procedures pursuant to chapter twenty (20) of the Code." This legislation clearly mandates teacher evaluation, with determination of criteria a board prerogative but procedures a matter for collective bargaining.

Purpose

This study will attempt to determine what criteria and procedures are being used in the evaluation of teachers as perceived by those directly involved, namely classroom teachers and building level administrators. The study will also discuss the alternatives available, and will

attempt to show how the pieces can fit together into a total evaluation picture.

Statement of the Problem

Several questions have been raised about teacher evaluation in Iowa. The following six general issues will serve as a guide for this study:

1. What evaluation practices, if any, are followed in the state?
2. Is there general agreement across districts regarding the purpose, criteria, and procedures which are used in teacher evaluation?
3. Is there general agreement within districts regarding the purpose, criteria, and procedures which are used in teacher evaluation?
4. Is there general agreement between teachers and administrators in their perceptions of evaluation practices?
5. Is there general agreement between elementary and secondary teachers in their perceptions of evaluation practices?
6. Are teachers and administrators generally satisfied with current evaluation practices?

Definitions

Teacher evaluation - Any process which results in making decisions about teachers; included are both data collection and appraisal of the data.

Evaluation procedures - Any of a wide variety of means for collecting and reporting the data used in teacher evaluation.

Evaluative criteria - Any of a wide variety of items which are or can be used as standards for judging the ability of a teacher.

School district - In this study the term will generally be restricted to public school districts in the state of Iowa, with specific application to the larger districts in the state.

Delimitations

The information to be gathered in this study will come from a sample of the larger school districts in the state. It has been assumed that the larger districts generally set the pattern for the other districts in the state in most major policies, so that the practices observed should be a fair representation of state-wide evaluation practices. The districts themselves will be voluntarily involved in the study and because of their openness may not be representative of all districts in regard to their evaluation program. The participants will also be volunteers, and since those who have strong personal biases would appear to be more willing to respond, the results may show more extremism than would otherwise be expected. Finally, since little work has been done on this topic, a lengthy questionnaire is planned which will cover many topics lightly. This in itself may cause some to not respond, making many conclusions tentative at best.

CHAPTER II. LITERATURE REVIEW

History of Teacher Evaluation

> When teacher evaluation is discussed, contemplated, or written about, there is always agreement on one point - the need for it. However, opinions on why the need exists are as divergent as the groups that have the differing views. To further complicate this issue, opinions differ and often conflict on the goals and purposes of evaluation, on the criteria for evaluation, how it should be done, who should do it, and even on whether it is possible to have meaningful evaluation (Wicks 1973).

Teacher evaluation has probably been around in one form or another for as long as there have been teachers. Evaluators include students, principals, school board members, parents, and members of the community who work with the products of the educational enterprise. Many of these evaluations are based upon distorted sampling of a teacher's behavior, but they are nonetheless evaluations, i.e. determinations of the worth of the teacher.

Teacher evaluation as we recognize it today had its roots in industry with its movement to scientific management as brought to Bethlehem Steel Company by Frederick Winslow Tyler (Davis 1964, p. 43). By 1912, the measurement movement had been recognized by the National Education Association, and the use of rating forms in research on teaching occurred as early as 1915 (Davis 1964, p. 45). Following World War II there was renewed interest in teacher effectiveness as a result of evaluating military instructional programs (Davis 1964, p. 52).

Hubert Wilbur Nutt (1920) in his book The Supervision of Instruction gave a suggested outline for teacher rating. He mentioned the following nine items to be considered when evaluating teachers:

1. Intellectual ability: measures of native intelligence as shown by tests and practice.
2. General scholarship.
3. Special scholarship.
4. Ability to express thoughts: both in use of language and speech patterns.
5. Teacher ability:
   a. Mastery of principles and methods.
   b. Intelligence and resourcefulness in selecting and adapting devices.
   c. Definiteness of lesson planning and skill in following plans.
   d. Skill and reliability of technique, checked by class visits.
   e. Ability to secure desired results, which includes study habits as well as subject matter comprehension.
   f. Ability to test and grade definitely and accurately the work of pupils.
6. A philosophy of school discipline.
7. Personal appearance (which Nutt admits is only a matter of personal opinion).
8. Qualities of leadership such as assuming responsibilities and taking initiative.
9. Professional attitudes: involvement in and support for school and intellectual pursuits, public interest, etc. (Nutt 1920).

With only a cursory study of the evaluation proactices used today, it is possible to conclude one of two things; either Nutt was considerably ahead of his time, or we have not seen much change in teacher evaluation since 1920.

In more recent times the topic of teacher evaluation has become a public issue. A combination of rising property taxes which were connected to rising costs of education, and increased militancy on the behalf of teachers, has led to a call for economic accountability for the schools. Since the largest portion of most school district budgets is due to professional salaries and fringe benefits, the focus has often become teacher accountability. In addition, with the passing of the Post War Baby Boom generation, declining enrollments have forced the reduction of teaching staffs. As a result of these and other forces, several states have enacted legislation which requires teacher evaluation.

Perhaps the most conspicuous piece of legislation was passed in California in 1971. The Stull Act required that each school board establish standards of expected student growth and indicate how it was to be assessed. Staff competence was then to be determined by attainment of the standards of expected student progress. In addition, teachers were to be assessed on the basis of their performance of other duties and on their effectiveness in maintaining proper control and preserving of a suitable learning environment (Redfern 1973).

Another example of an evaluation law came from Kansas. This law required that all certified staff be evaluated, and that the evaluation program have both teacher and community involvement. The law required that there be a clear delineation of who is to evaluate whom, how, when, and to what end (Sarthory 1973). Several other states also require teacher evaluation. At this time Oregon, Florida, Washington, and Hawaii require that the evaluation be at least annually (Redfern (1973). Connecticut requires that the evaluation of personnel be on a continuous basis (Saif 1974). Additional states are following with a wide variety of teacher evaluation laws. As previously mentioned, Iowa currently has mandated teacher evaluation under 279.13 of the Iowa Code. Whether any of this has had an effect on teacher evaluation remains to be seen. Whether it has led to the improvement of instruction is clearly in question.

Teacher Evaluation Defined

In order to discuss teacher evaluation, it is important that we agree upon a basic definition. Saif has said that "(e)valuation is a

collection of data for further decision making" (Saif 1974, p. 2). This may be thought of as a three phase process that involves selecting the attributes desired in the teacher, describing the teacher in terms of these attributes, and arriving at a judgment as to the merit or worth of the teacher (Howsam 1973). Note that the requirement is clearly twofold: collecting data and making judgments, in that order. The Stull Act made evaluation of teachers far more specific by prescribing exactly what things were to be considered. The crux of the issue still must lie in the selection of the standards by which the judgments are to be made, and in detailing the procedures for gathering data.

When speaking of teacher evaluation, it is important that we realize that teacher evaluation is more than just a look at individual teachers. It is, or should be, a systematic way to look at a complete school district program and practices. The National School Boards Association (1971) listed ten characteristics of an effective teacher evaluation program:

1. Board policy should view evaluation as a means for improving instruction.
2. All procedures, forms, job descriptions, and criteria should be developed cooperatively between the administration and staff.
3. Criteria should be explicit, objective, and relate to behavior which is directly related to student performance and advancement of the instructional program.
4. Evaluation is a continuing process and is both formal and informal.
5. A variety of techniques should be used for assessing performance.
6. The program should encourage continuing self-evaluation and improvement.
7. Observation includes follow-up at which the teacher is given a signed copy of any written evaluation.
8. The teacher has a right of appeal of any unfavorable evaluation to the superintendent and the Board.
9. Evaluators are to be trained in evaluation.
10. The program should include reliable measures for the evaluation of the performance of evaluators (National Association of School Boards 1971).

It should be noted that an evaluation system must begin with the specification of the goals or purposes of evaluation and then identify those variables of a teacher's behavior which fit with the prescribed goals (Feldhusen 1976). Often these goals can be determined at a pre-evaluation conference where purposes, goals, and methods can be freely discussed. The actual evaluation ought to include two or more observations by two or more evaluators. Following the observation there should be a post-observation conference which includes the right of rebuttal and appeal (Wicks 1973). McKenna has stated that "school staffs can be held accountable only to the extent that conditions are present under which competence can be demonstrated" (McKenna 1973a, p. 17). This can best occur when the teacher is a partner in the process, which includes partnership in determining the procedures and criteria. It has also been recommended that evaluation take a balanced, humane view of everything the teacher does which effects students. This would require checking a large number of items, but realizing that no teacher can be expected to score well on all of them. It should also be recognized that many things can happen as the result of an evaluation, and not all of them are just to the teacher. A good evaluation program helps improve teachers, may identify teachers with special abilities, and may reflect needed changes in administration or available facilities (National Association of School Boards 1971).

Prerequisites of a Good Evaluation Program

> Evaluation is not an end in itself. If a program of evaluation creates anxiety that interferes with good teaching, if it stimulates or reinforces hostility, if it simply takes so

> much time from learning that the net gain is negative, let's forget it (McKeachie 1970, p. 7).

In order for us to understand what is meant by a good evaluation program, it is perhaps instructive to contrast it with what is not meant by a good evaluation program.

> Staff evaluation that is well-planned and clearly understood can be a tremendous asset in the positive upgrading of a total local school district's staff performance; an evaluation that is unilaterally decided upon, poorly communicated as to the underlying rationale or not clearly understood by either the evaluator or the evaluatee can become one of the most destructive elements ever introduced into a local school district (Herman 1973, p. 30).

Harold J. McNally has given us eleven characteristics of what he considers to be a good evaluation program.

1. Purposes are clearly stated in writing and well-known to both evaluators and those being evaluated.
2. Policies and procedures should reflect knowledge of research related to teacher evaluation.
3. Teachers know and understand the criteria by which they are evaluated.
4. The evaluation program is cooperatively planned, carried out, and evaluated by teachers, supervisors, and administrators.
5. The evaluations are as valid and reliable as possible, where valid means important to the learning of children, requires adequate sampling of behavior, and exhibits criteria which are related to the needs and conditions of the local setting; and reliability means that evaluators agree, being aided by guidelines, having training, and limits on the range of criteria used per visit.
6. Evaluation is more diagnostic than judgmental.
7. Self-evaluation is an important objective of the program.
8. Self-image and self-respect of the teacher is maintained and enlarged.
9. Evaluation encourages creativity and experimentation in planning and guiding the teaching-learning experiences provided children.
10. The program makes provision for clear, personalized, constructive feedback.
11. The program is seen as an integral part of the leadership role of the principal and of the program of in-service (McNally 1973).

While a program need not exhibit all these characteristics, such features do appear to be worthy goals.

Another requirement for a good evaluation program is objectivity of the ratings by observers. This is included in McNally's reliability and validity, but merits a closer look since this is one area where a great deal of friction can occur between staff and administration. As Ryans has said:

> No evaluation system is objective if it is primarily dependent on intuitively derived assessments. A school administrator who visits a teacher's classroom on one or two occasions during the year, then bolsters such impressionistic data with personal perceptions of the teacher's cooperativeness and general intelligence, could hardly be said to have engaged in an objective evaluation of the teacher (Ryans 1960, p. 14).

Systematic evaluation is necessary as has been shown by a wide variety of court cases at both state and federal levels which have dealt with nonrenewal. Cases involving race, lack of due process, violation of freedom of speech, etc., could have been avoided had there been an evaluation system which had been effectively carried out (DeVaughn 1971). Administrative due process does not appear to be difficult if a few guidelines are followed. First, the evaluation program must identify strengths and weaknesses early in the teacher's employment. Second, the individual should be involved in self-assessment. Third, it is necessary to provide supervisory assistance to bring performance up to acceptable levels if possible. Finally, specific reasons must be given if performance is judged to be inadequate (DeVaughn 1971). If these four steps are followed and carefully documented, many of the problems with due process can be avoided.

Redfern has made the following observation:

> The need for an effective evaluation program is one thing: the means of achieving workable practices is another. The crux of the problem lies in the development of techniques which will be satisfactory and acceptable both to the individual being evaluated and those responsible for the evaluating (Redfern 1972, p. 10).

One key to upgrading evaluation programs is knowledge of research. "School systems vary considerably in their practices. Generally they are not implementing recent research findings" (Egnatoff ca. 1974, p. 20). Some trends appear to be emerging. Schools are moving from spasmodic to planned and continuous evaluations, from superimposed to cooperatively planned and implemented programs, and from supervisor staff evaluations to more staff self-evaluations (Egnatoff ca. 1974).

How acceptable these changes will be to teachers remains to be seen. Herman suggests that "the degree of acceptance of any finalized system of staff evaluation will be greatly dependent upon the quality and quantity of communication provided during the process of evolution of the evaluation program" (Herman 1973, p. 19). One means of helping make this more positive could be to "develop new ways to reward the professional performance of teachers" (Pedersen 1975, p. 20). It is clear that the only path to success involves teamwork.

Purposes of Evaluation

"The purposes of teacher evaluation should grow out of clearly stated goals of the school system and should contribute to the accomplishment of these goals" (National Center for Educational Communication 21-C, 1971, p. 5). Several possible purposes have been suggested, but the following

list from the Teacher Evaluation PREP-21 of the National Center for Educational Communication seems to contain the most commonly cited:

1. Improving teaching
2. Rewarding superior performance
3. Supplying information to modify assignments (which includes dismissal)
4. Legal protection for individuals and the school system
5. Validation of selection processes
6. Providing a basis for career planning and individual development (including advanced degrees and inservice) (National Center for Educational Communication C-21 1971).

The following quote from Paul Dressel (1976) sums up these purposes:

> Instructors, interested in improving teaching competence, need to define the learning expected, the stages of student development, and the means of motivation. They require continuing and detailed information to suggest means of improvement. They are likely to seek insights from their students and generally welcome assistance from learning and evaluation services. Clearly, this evaluation is formative. Instructors have little need for or interest in comparative data on other instructors.
>
> Administrators, faced with a variety of personnel and budgeting decisions, want a brief evaluative summary and usually prefer comparative evidence to guide their decisions. Such information is however often not reliable, valid, or adequate (Dressel 1976, p. 337).

Apparently, the kind of evaluation performed depends a great deal on the purpose of the evaluation.

The first purpose of evaluation listed above is "improving teaching." Medley has buttressed this by saying: "Unless a program of teacher evaluation improves the instruction in a school, it has no reason for existing" (Medley 1973, p. 33). Carlson and Park (1976) claim that for this improvement to occur it is necessary to improve teaching behavior, to clarify needs, goals, and objectives, and to otherwise make instructional programs more effective by reallocating resources and positively reinforcing good teaching while negatively reinforcing bad teaching.

This would include being able to eliminate the "worst" teachers according to some predetermined criteria. Others contend that "the question 'Is he or she a good teacher?' is meaningless and should be replaced by the question 'What is this teacher good for?'" (Krasno 1972, p. 4). Regardless of the choice of position, it would seem that "if the goal is to improve instruction, then improvement should be valued, encouraged, and rewarded by the system" (Medley 1973, p. 34). However, a spokesman for the American Federation of Teachers has said, "We are agreed that the salary of a teacher must not be tied to his pedigogical worth" (Seldon 1969, p. 1). The question remains how to reward teachers for superior performance without using monetary measures which are so unacceptable to the unionists, while still offering something which will truly serve as an incentive. The literature yields no answer at this time.

One important piece of information regarding the purpose of evaluation came from a study of twenty Iowa school districts. It was found that teachers who felt that evaluation was for improvement of instruction were generally supportive of evaluation, while those who felt that evaluation was for administrative purposes (items 2-6 above) tended to regard evaluation in a negative manner (Zelenak 1973). One problem does arise.

> The reason for appraisal is often said to be 'to improve instruction,' but the methods seldom relate to instructional practice and even less often to the results of instruction. As typically conducted, faculty evaluation cannot be seen as a way to improve instruction (Cohen and Brewer 1969, p. 52).

A second question which will be dealt with in more detail later concerns how evaluation should be implemented if improvement is the key. That is, should we be most concerned with the behavior of the teacher or should

we be more concerned with students. As Pine and Boy have stated, "The ultimate purpose of effective evaluation is the growth and development of the student" (Pine and Boy 1975, p. 19).

The second general purpose of teacher evaluation revolves around its use as a tool for administrative decision making. "One positive by-product of a sound evaluation program is the production, recording and storage of masses of information that can be drawn upon at future times for the purpose of making knowledgeable administrative decisions" (Herman 1973, p. 34). There is some question among educators as to whether information gathered to improve instruction can and should be used in administrative processes including nonrenewal. Perhaps more important is the fact that "(s)upervision of teacher performances and measurement of consequence data in school settings will need to be improved so that retention, dismissal, or promotion of teachers is based upon the teacher's ability to achieve intended consequences" (Kelley 1974, p. 26). Evaluation for administrative purposes must possess several characteristics not necessarily required of evaluation for improvement of instruction. First, it must be "fair" in the sense that a teacher must not be penalized for teaching conditions over which he has no control (Gage 1972). Also, the evaluation process should not be used to enforce discipline and conformity within the school, especially when these things might not be indicators of the teacher's performance (Seldon 1969).

In summary,

> (i)f teacher evaluation is used for the positive improvement of the teaching process, it will be worthwhile. If it is used in a punitive and negative manner, it will not result in improved instruction. In fact, the more threatening it becomes, the more obstacles it creates to teacher effectiveness (Wicks 1973, p. 42).

Who Should be Involved in Evaluation

Who should determine evaluation practices? Who should be involved in evaluation procedures? Who should serve as the evaluator? These questions are at the heart of determining the procedures to be used in an evaluation program. Redfern says,

> The impetus for instituting or revising an evaluative program should come from the highest level. The superintendent must believe genuinely in the idea. The board of education must understand fully what is involved, be willing to give the program status and policy-level support and authorize the financial support needed to carry out the program (Redfern 1972, p. 9).

Feldman claims that

> experienced classroom teachers must be involved fully in the development of any program for evaluation of professionals, or of schools, or for changing teacher training programs. First, because they have a great deal to offer; second, because their very involvement will lead in the process to improvement of their own teaching (Feldman 1972, p. 4).

It has been noted that the involvement of teachers in planning evaluation procedures yields better understanding of these procedures than when they are unilaterally determined (National Center for Educational Communication 1971).

Herman (1973) lists seven possible evaluators: peers, students, self, lay residents, outside paid consultants, central office administrators, and immediate supervisors. It is clear that several of these are involved in evaluation whether we ask them for their opinions or not. Saif (1974) suggests that the evaluator be someone in the system who can help and support the teacher. Wicks (1973) suggests that students' ratings, voluntary self-appraisal using ratings, audio tape, or video tape, or mutual self-appraisal with peers offer possibilities. In a 1955-56

study by the National Education Association, it was found that in 37 per cent of the districts surveyed, the principal was the only evaluator, with joint ratings by the principal and a supervisor accounting for 27 per cent of the responses and separate ratings by principal and a supervisor occurring 17 per cent of the time. It was said at that time that "the principal is clearly the primary rating officer" (Davis 1964, p. 57). There is nothing to indicate that this has changed after more than twenty years. The question of outside evaluators arises when we consider whether we wish to use formative or summative evaluation. "Most of us do not want to be evaluated for an outside audience.... If we mean to do summative evaluation of instruction - having an outside audience - we should do it with external evaluators; and we should set up the ground rules and purposes in a clear-cut, understandable way" (Hastings 1973, p. 141). Hastings went on to point out that summative evaluation is not for reformation because it is necessarily too general and less concerned with small units which can be modified for improvement.

Problems With Evaluation

> Most appraisal procedures and instruments have been inadequate and highly subjective and have been administered under an assumption that the superior somehow possesses the required competencies to make the correct judgment, usually without involvement of the evaluatee in the process through self-appraisal, when the evaluatee perhaps best knows his strengths and weaknesses and could adequately state his professional need for help if invited to do so in an open, relatively threat-free climate (DeVaughn 1971, p. 2).

Teacher evaluation has been fraught with problems for years. In some systems, both teachers and administrators regard evaluation as a required ritual and nothing else. "When evaluation is only a formality,

it is of little value to anyone, and is likely to do more harm through attaching negative labels instead of giving constructive advice where needed" (Davis 1964, p. 63). Drumheller (1974) contends that most evaluation consists of picking out the "socially sophisticated 'nice guy'" rather than focusing on the real crucial characteristics of maintaining an orderly classroom and serving as a learning facilitator.

Several problems have been listed by Herman (1973). First, Herman claims that administrators have often neglected to tell an employee what exactly was expected of him, thus holding him responsible for assignments of which he was unaware. Secondly, Herman claims that administrators forget to meet with the employee to identify what is to be achieved. Thirdly, Herman has charged that the goals teachers are expected to meet are not stated in terms which can be measured for evaluative purposes. Finally, Herman notes that even after weaknesses have been pointed out in a teacher's performance, administrators often do not take the responsibility of assisting the teacher in the elimination of the weaknesses through in-service and job upgrading and are not consistent in a follow-up program to insure improvement.

Another problem involves the cost of a good evaluation program. Roth and Mahoney (1975) have suggested that if costs in terms of time, money, and effort cannot be justified in terms of the information gained, then alternative sources of information should be considered. However, in a recent study (Shea 1977) which involved Kansas junior high schools, it was noted that principals were the primary evaluators and they spent only one to three hours per week on teacher evaluation. This should bring into question the priority which has been given to teacher evaluation, especially

in a state where evaluation is mandatory (Sarthory 1973).

Perhaps the most haunting problem of the entire teacher evaluation movement is one of measurement. Prejudice, inconsistency, subjectivity, allowing personality to enter in, measuring too many elements, continuing prior viewpoints (halo effect), and consistent over or under evaluation are among the many difficulties (National Center for Educational Communication 1971). "The most obvious fact about the measurement of teacher behavior is the lack of universal agreement about what is to be measured" (McDonald 1972, p. 61). There is a need to find some connection between what is measured and the products of education. "We absolutely must pin down the connections between the inputs and the outputs of education; without that kind of theoretical structure we can flounder indefinitely in our efforts to improve the process" (Mood 1970, p. 9). One of the additional difficulties brought about by this lack of connection between input and output is due to a complex interaction which takes place because "the teaching act varies from person to person, and from situation to situation" (Barr 1968a, p. 5). Popham has stated, "The history of teacher effectiveness research is replete with failure after failure in efforts to devise defensible measures of how well a teacher teaches" (Popham 1973a, p. 35).

Some have suggested that the real importance of teaching can only be found by considering the long term effects of the teacher on his pupils. Unfortunately,

> (t)he long term impact of any given teacher is extremely difficult to assess because of the 'contaminating' involvement of other people and other educational programs. Furthermore, the immediate effect may be unobservable or have limited meaning in relation to the general overall educational goals of a school system (Pedersen 1975, p. 14).

In spite of the difficulties, it is important that research continue. Bolen (1973a) has encouraged educators to continue in their efforts to learn as much as possible about what goes on in the classroom. Unfortunately, in the meantime many evaluations will still be a ritual which does nothing to improve the instructional atmosphere of a school. "These teacher evaluation schemes seem to be predicated on the following proposition: No Teacher a Loser (author's emphasis)" (Popham 1975b, p. 284).

Attitudes Toward Evaluation

Robert L. Wolf has made what is perhaps the greatest understatement in teacher evaluation literature: "Teachers are not fond of evaluation" (Wolf 1973, p. 160). Research has shown that administrators have more positive feelings about teacher evaluation than do teachers (Reavis 1975), and attitude does have an effect on evaluation. "Several researchers have found that the extent of the benefits a teacher receives from evaluation is related to his attitude toward evaluation" (National Center for Educational Communication 1971 21-D, p. 2). For many teachers, the benefits are difficult to determine.

> Ultimately, teachers view evaluation as a unilateral exercise of power. The absence of any similar evaluation, not only of the evaluator as evaluator but as administrator, presupposes that teachers are the only ones in need of constant improvement. However, the almost punitive nature of evaluation is seen as a threat to professional status and personal freedom. It is regarded as a form of trial rather than an instrument for improving classroom effectiveness (Pedersen 1975, p. 17).

Sometimes there are problems with the criteria used for evaluation. As Wolf has said:

> Teachers probably believe that the standards for evaluating

> what is effective teaching are too vague and ambiguous to be worth anything. They feel that current appraisal techniques fall short of collecting information that accurately characterizes their performance. They perceive the ultimate rating as depending more on the idiosyncrasies of the rater than on their own behavior in the classroom (Wolf 1973, p. 160).

Barbara Reimers, president of the National School Board Association in 1974 stated,

> If some teachers view evaluation as a threat, a way for administrators to vent personal dislikes, or for school boards to get rid of high-salaried older teachers in exchange for less expensive beginners, it probably is because some school districts have yet to discover even a concept, much less a method, of evaluation that assures the school board of measurable productivity and the teacher of fair treatment and advancement based on achievement (Pierce and Smith 1974, p. 8).

Mallery continues this thought by saying,

> Until teacher-evaluation programs do pay attention to strengths, and to building on them toward the kinds of professional growth and soaring that ideally we would like to have for our teachers, 'evaluation' will keep right on being something that is inflicted on teachers. Such an exposure of faults, under unpleasant conditions, is almost a professional assault, one that invites defensiveness and divisiveness in a faculty (Mallery 1975, p. 3).

What can be done? Some have said that we must lean upon research findings. "(T)he distressing truth is that systems for assessment and evaluation of teacher performance must be erected on technical foundations which more closely approximate balsa wood than concrete" (Popham 1971a, pp. 11-12).

Evaluative Criteria: General Comments

> The question of standards is one which plagues all evaluation efforts, Peter W. Airisian, 1974 (Roth and Mahoney 1975, p. 7).

The entire weight of teacher evaluation balances on the importance

of having usable and justifiable standards for making judgments regarding teacher effectiveness. Cynicism is not uncommon in discussions of criteria as the following pair of quotations will show:

> A supervisor of student teachers in elementary education remarked that a most essential trait for a young teacher is probably the ability to locate and assume her place in the pecking order of the group of teachers and administrative personnel in which she finds herself (Sorenson and Gross 1967, p. 6).

> "... the real working criterion of teaching success is the ability to secure and to hold a position. Having said... this, and in spite of the lack of evidence for a general trait which can be called teaching ability, supervisors and placement officers seem to be confident that they can tell a good teacher when they see one (Worcester 1968, p. 124).

Cynicism aside, as early as 1920 Nutt reported:

> (t)he rating of the teacher... requires a most careful study and accurate analysis of the teacher's classroom performances, in light of a definite set of standards. Vague generalizations and broad guesses must give way to scientific analysis and accurate measurement. If a set of definite standards cannot be set up, by means of which the efficiency of the teacher can be reliably measured, then the rating of teachers should be abandoned altogether (Nutt 1920, p. 215).

Bolen (1973b) has given six standards which evaluative criteria should possess: social validity, conceptualization of teaching behavior, stability over short time periods, variability within the population of teachers, relatedness to good teaching, and measurability. In addition, Hall (1974) has recommended that criteria be established <u>for a particular teacher</u> within the philosophy of the school.

There is considerable division as to whether teacher performance or student performance is the better indicator of teaching effectiveness (Saif 1976, Medley 1973, Rosenshine and Furst 1975, Bolen 1973a, Gaines 1973, Feldman 1972, O' Hanlon and Morteson 1977). Feldman has said,

> They (teachers) should be evaluated for what they know and for those parts of the educational totality over which they have control. They should be evaluated not on the basis of achievement scores of pupils - which involve many other factors besides teacher input - but on the basis of their own behaviors in the context of knowledge about which teacher behaviors or skills or competencies effect learning (Feldman 1972, p. 3).

Smith has supported this position by saying: "No one should be held responsible for outcomes beyond his influence. But the teacher can and ought to be held responsible for following tested methods of diagnosis and treatment" (Smith, 1972, p. 83). On the other hand, O'Hanlon and Morteson have stated, "Emphasis should be placed on evaluating student outcomes in the assessment of teacher performances. What the student learns is the prime criterion for the evaluation of teaching" (O'Hanlon and Morteson 1977, p. 3).

The middle ground between these opposing positions is not secure. Although some would ask that most criteria relate to student performance and to teacher activities which promote student learning (Saif 1974),

> (r)esearch has provided the practitioner with little evidence to assist him in identifying the cause-effect relationships in the teaching-learning situation. Consequently, since he cannot refer to research to tell him what effective teaching should be, the decisions he makes are based on intuitive, experienced-based judgments (Musella 1970, p. 21).

Sources of Evaluation Data

In order to have an evaluation program, it is clear that there must be some means for collecting evaluative data. This data can come from a wide variety of sources. An Ohio report based upon a survey of school personnel from superintendents through teachers in fifty Ohio school districts produced a list of nine different sources: observation,

rating scales, Redfern Model, self-appraisal instrument, micro teaching, Interaction Analysis, Nonverbal Teacher Behavior Category System, video taping, and the use of achievement tests (Commission of Public School Personnel Policies in Ohio 1972). Other sources have suggested the use of students, peers, support personnel, and even parents (Carlson and Park 1976). "It should be kept in mind that the best evaluation of any individual's performance is the balanced evaluation which draws upon numerous samples of behavior and numerous sources of information" (Carlson and Park 1976, p. 15).

The actual behavior monitored can also vary. It may include in-class behavior which can be observed by students, the teacher himself, peers, administrators, or even paraprofessionals. It might include out of class behavior which is also generally observed by the teacher himself, students, peers, administrators, aides, cooks, and custodians. Finally it is possible to view student accomplishment through teacher made tests, standardized tests, student self-report devices, observance of student behavior, and student products or projects (Boltan 1973b).

Generally, in order to keep observations focused it is necessary to use some type of instrument such as a scale, a check list, or a more complex analysis system. Regardless of the instrument used, it should possess certain recognizable characteristics. These include relevance to the situation, reliability and objectivity in usage, validity, fidelity in reflecting the actual performance, and ease of administration (Weisenstein 1976). "The staff and principal must be in agreement regarding all elements within the instrument, how it is to be used and for what purposes it is to be used" (Weisenstein 1976, p. 16)?

Implementing an Evaluation Program

Developing and implementing an evaluation program is not a process which can be done in haste. Speicher (1972) recommended the following series of steps for developing a teacher evaluation program:

1. Begin with a review of literature.
2. Involve teachers, administrators, and board members if possible.
   a. Review plans currently in use elsewhere.
   b. Develop goals.
   c. Identify a model of teaching effectiveness.
   d. Identify indicators of effectiveness.
   e. Determine procedures: who is to be evaluated by whom and how.
   f. Determine how data is to be collected: instruments, etc.
   g. Design a time schedule and steps in the program.
   h. Develop a decision making system.
   i. Design a system to give feedback on effectiveness of the evaluation plan.

Mallery (1975) suggested that a school try out a wide variety of interesting instruments, strategies, and approaches before making a final decision. Saif (1974) advised that teacher evaluation should be an ongoing process which gathers data on both strengths and weaknesses with appropriate measures taken to improve student achievement, and that teachers and evaluators must decide together what should be evaluated using what criteria and how the data will be gathered.

When complete, the evaluation program should be "(a) self-correcting system... which will identify errors and make changes in procedures before harmful effects occur" (National Center for Educational Communication 1971 21-A, p. 4). The key to a successful evaluation program can be summed up in one word: time. Herman (1973) suggested three and a half <u>years</u> be allowed for development and implementation of a plan for staff evaluation.

Teacher Effectiveness Research

> There are skills and attitudes which make for good teaching and most people can learn them (Laurits 1967, p. 33).

Perhaps the foremost name when discussing teaching effectiveness is David G. Ryans. His work, Characteristics of Teachers (1960), represented a large number of studies involving more than a decade of research. Ryans has said:

> To adequately carry out evaluation of teaching, we must have evaluative criteria that have been agreed upon, we must have some taxonomy and description of characteristics that comprise teaching behavior, and we must have some means of assessing or measuring those characteristics (Ryans 1967, p. 50).

Unfortunately, a few years later Spencer and Boyd reported,

> Education researchers report significant progress within the past decade in attempting to link what teachers do in classrooms to children's educational progress. Even so, the connection is by no means understood. For many valid reasons, educators are still unable to make a direct connection between desirable student outcomes and specific acts in the teaching process (Spencer and Boyd 1977, p. 679).

Rosenshine and Furst have summarized the results of much of this research. "The five variables which have yielded the strongest relationships with measures of student achievement are: clarity, variability, enthusiasm, task orientation and/or businesslike behavior, and student opportunity to learn (Rosenshine and Furst 1975, p. 54). On the other hand, a whole grocery list of suggested variables have not shown a consistent relationship with achievement. These include praise, warmth, indirect/direct ratio, flexibility, amount of teach/student talk, student participation, student or teacher absences, teacher time spent in class preparation, teacher experience, or teacher knowledge of subject area (Rosenshine and

Furst 1975).

Before we decide to ignore all the work done by those studying teacher effectiveness, let us look a bit deeper into what has been done, both good and bad. Quite frankly,

> (r)esearch in the behavioral sciences, and particularly in teacher education, has been extensive, yet powerfully integrated studies rigorously applied have not demonstrated strong relationships between specific teacher competencies and student outcomes (Houstan and Jones 1974, p. 23).

There are good reasons why this should not be surprising.

> At the beginning, research concentrated on the teacher's characteristics, personality, method, or behavior; and unfortunately, the results were negligible. This is expected, because such an approach falsely assumes that teaching effectiveness is determined by one variable in the teaching-learning situation, the teacher (Saadeh 1970, p. 79).

In addition to this rather shaky assumption, Berliner (1975) has claimed that there has not been enough data collected about the individual differences among students to see if different teaching behaviors have different effects on different types of students. Another problem rests in the general stability of teacher characteristics. "It appears that teachers do not, by and large, remain in a stable ordering on measures of teacher effectiveness" (Berliner 1975, p. 23). Berliner went on to report that the correlation of two or more measures of teacher effectiveness often have an average correlation of an 0.30. Part of this difficulty involves the actual sampling of behavior. The mere absence of an item used in an analysis may not mean a lack of it (Roth and Mahoney 1975).

The Competency Movement

In the early 1970's a new movement surfaced in teacher education: Competency Based Teacher Education. This movement contrasted with the effectiveness proponents of the previous decades. Dropkin (1973) illustrated the differences by noting that effectiveness implies that someone is adequate to yield the intended result, thus implying a casual relationship between teaching and learning with the focus on changes in learner behavior. Competence on the other hand has two components: knowledge and ability. These are gauged against specific criteria of teacher behavior. Dropkin further noted that effectiveness depends on competence. This is further illustrated by Powell who said: "... we need a theory which allows us to describe what we mean by competent teaching in terms of what teachers actually do independently of what their pupils do" (Powell 1970, p. 138). It should be clear at this point that many see effectiveness as determined by pupil progress, while competence is determined by the teacher's ability to cope with a certain class of problems that a teacher encounters on the job (Coker and Coker 1975). Unfortunately, this use of language is not quite so clear as Biddle has written: "In its broadest meaning, competence is an individual's ability to produce agreed-upon results" (Biddle 1964, p. 2).

Assessment of Effectiveness and Competence

> The greatest problem in predicting teacher effectiveness lies in the definition and evaluation of teacher effectiveness.... First there is a need of a realistic definition of performances that achieve desired goals, and second, we need accurate and unbiased means of measuring these performances (Abell 1968, p. 49).

> The measurement of teacher effectiveness is no easy task.
>
> ...(T)wo classes of variables are minimally necessary in the study of effectiveness: teacher behaviors (an independent variable) and teacher effects (a dependent variable). The problem becomes complex because teacher-pupil interaction is imbedded in historical, social, and physical context that constrain and interact with it (Biddle 1964, p. 5).

One of the major problems in effectiveness research has been the use of administrators and supervisors as judges of effective teaching. "In almost all of the studies, it has been assumed that a person with a certain title is a competent judge. It is amazing that this assumption has not been more seriously challenged" (Worchester 1968, p. 124). A study by Tolor (1973) attempted to ascertain the degree to which those who are concerned with education can agree upon who is regarded as a highly effective teacher or who is low in effectiveness. Four groups (administrators, faculty, students, and parents) were asked to identify high and low effectiveness teachers. Highest agreement was found between administrators and faculty on both high and low effectiveness. Very little agreement was noted between students and any of the other groups, which indicated that students apparently use different standards.

Another problem has been in the instability of those factors which are considered to be indicators of effectiveness.

> In general, the term "effective teacher" has been taken to mean that a teacher remains effective across a number of years. Yet on the basis of these (five) studies, evidence on the consistency of teacher effects is weak because correlations as high as .5 were obtained in only one study,... and all other correlations were about .35 or much lower. There is a need for further research to establish whether terms such as 'effective teaching' or 'ineffective teaching' have any stable meaning (Rosenshine 1970, p. 650-651).

To add to the frustration, Flanders has noted: "It is easier to identify

poor teaching than it is to identify superior teaching. The characteristics of the former are more consistent than the latter" (Flanders 1964, p. 227).

The assessment of competence is seen by many educators as a different matter. Competence is described as comparing the teacher with an "ideal teacher," whose attributes are described in terms of what that "ideal" does (Roth and Mahoney 1975, p. 21). Barbara Reimers, president of the National School Board Association in 1974, has stated:

> Factors of competence and incompetence must be identified if teacher evaluation is to mean anything at all.... I do not, in any case, subscribe to the notion that because factors of competence or incompetence are difficult to identify, they don't exist (Pierce and Smith 1974, p. 10).

The major problem is the link between the competencies and the results of the application of these skills in the teaching of children (Rosenshine and Marten 1974). "It does not seem likely that an adequate method for evaluating teacher competence will satisfy educators, social scientists, and the public" (Worcester 1968, p. 263).

Resnick and Reinert (1973) have a set of twelve criteria which can be measured by using classroom samples which they claim will reveal a teacher's competence. Bain, Billings, and Houston (1973) have published a different list of thirty-three competencies. Other sources list hundreds of competencies, yet the question remains: What are the crucial competencies? We simply do not have a definitive list (House 1975). Catalogues of competencies exist which have been called "an indefinite portion of an undefined set of vague but tediously classified and cross classified homilies" (Maxwell 1974, p. 308). No attempt has been made to show them to be independent, mutually exclusive, or exhaustive.

Competencies seem to have been identified by a consensus of "experts" who agree that they truly have a competency (Maxwell 1974, Rosener and Kay 1974). Rosenshine warns: "Teaching competencies are not matters that can be decided by a poll of concerned citizens or be legislated by state legislatures" (Rosenshine 1973, p. 28).

Another difficulty arises in attempting to make the connection between competencies and teaching behavior. "The hypothesis is that if a teacher can articulate knowledge, skills, and attitudes thought to be important and related to competencies then it is more likely that this teacher would exhibit the appropriate behavior" (Coker and Coker 1975, p. 10). As McDonald has said: "How do you know that the teacher who acquires the competencies for which you are training them will be a more effective teacher" (McDonald 1974, p. 296)? Maxwell (1974) has concluded that it lies with the proponents of a competency approach to teacher education and evaluation to show that there exists a discrete set of competencies which can be measured and which define the successful teacher.

Effectiveness Research

The material available on teacher effectiveness can be roughly divided into three categories: personal characteristics, classroom behavior, and interpersonal relations.

> Although there is considerable evidence that particular personality characteristics of teachers have a discernible influence on pupil behavior, there is little evidence that certain personality characteristics are more desirable than others for teaching in general (Musella 1970, p. 17).

B.O. Smith has also commented in this regard:

> Knowledge of the characteristics of teachers is of limited

> value. If the characteristics are basic elements of personality, it is doubtful that knowledge of their influence on teaching behavior would be of value anyway, because such factors are not subject to serious modification by pedagogical modes of influence (Smith 1973, p. 72).

Research has produced a variety of traits which have been found to distinguish effective teachers. These include flexibility, knowledge of subject, informality of style, verbal fluency, enthusiasm, clarity, warmth, task orientation/business-like manner, praise, and sense of humor (Table 1). Of those listed, "teacher enthusiasm may be the most powerful personality characteristic of all when it comes to effective teaching" (Hamachek 1975b, p. 303).

Table 1. Personal characteristics as criteria for effectiveness: citations

Characteristics	Generally indicated as criteria	Correlated with pupil achievement	Correlated with pupil affective growth
Flexibility/ variability	Hamachek 1975a Hamachek 1975b Glass 1974 Medley 1971	McKenna 1973a Brophy 1974	
Knowledge of subject	Hamachek 1975b Miller 1972		
Informal style	Hamachek 1975b		
Verbal fluency	Kulik and Kulik 1974		
Enthusiasm	Kulik and Kulik 1974 Gage 1972 Hoyt 1969 Medley 1971 Hamachek 1975b	McKenna 1973a Brophy 1974 Rosenshine 1975	
Clarity	Hoyt 1969 Glass 1974 Medley 1971	McKenna 1973a	

Table 1 (continued)

Characteristics	Generally indicated as criteria	Correlated with pupil achievement	Correlated with pupil affective growth
Warmth	Gage 1972 Hamachek 1975b		Hamachek 1975b
Task oriented/ business-like	Glass 1974 Medley 1971	Brophy 1974	
Praise	Medley 1971		
Sense of humor	Hamachek 1975a		

In the area of classroom behavior, another group of variables have been identified, including: skill in questioning, being well-prepared, indirectness, organization, use of discussion, and using structuring comments (Table 2).

Table 2. Classroom behaviors as evaluative criteria: citations

Behaviors	Generally indicated as criteria
Skillful in questioning	Hamachek 1975b, Medley 1971
Well-prepared	Miller 1972
High level (cognitive) testing	Miller 1972
Indirectness	Gage 1972
Organization	Gage 1972, Medley 1971
Use of discussion	Johnson and Radebaugh 1969, Medley 1971
Structuring comments	Medley 1971

With regard to interpersonal relations, the effective teacher is perceived to view the world from the student's viewpoint, to treat students with respect, to motivate students to do their best, to show interest in

pupils, and to have knowledge of pupils (Table 3).

Table 3. Interpersonal relations as evaluative criteria: citations

Qualities	Generally indicated as criteria	Correlated with pupil achievement	Correlated with pupil affective growth
Perceives world from student viewpoint	Hamachek 1975b		
Treats students with respect	Miller 1972		
Motivates student to do his best	Miller 1972		
Shows interest in pupils	Hamachek 1975b		Hamachek 1975b Peronto 1968
Has knowledge of pupils		Hamachek 1975b Peronto 1968	

> Despite the hundreds of variables that have been researched, we do not know how many more may be operating. Moreover, we have no way of knowing which variables are relevant until we have a notion of good teaching. The criteria that have been used have been derived from administrators' and supervisors' notions of good teaching, and so the question is begged rather than answered (Broudy 1969, p. 584).

P.M. Symonds summed up the problems as follows:

> I have seen successful teachers with loud, harsh voices, and also with very soft, indistinct voices. I have seen successful teachers who were lax, easy going, highly permissive and others who were strict and restrictive. I have seen successful teachers who were effusive in giving praise, but I have also seen successful teachers who never seemed wholly satisfied with what the children in their classes do (Zax 1971, p. 287).

In contrast to the scholarly research cited above, Chrisman noted that "a variety of criteria are used to identify superior teachers, but

the most frequently found was a teacher's willingness to assume extra duties (italics mine)" (Zax 1971, p. 285).

While the studies which yielded the results reported here have indicated some things which are considered to be indicators of effectiveness, other authors have either countered the claims or have listed things which do not distinguish effective teachers. For example, Johnson and Radebaugh have found that marital status, undergraduate grade point average, frequency of use of media, extent of television watching, reading of journals and magazines, or the teacher's judgment of his own sense of humor are not important items in determining whether a teacher is effective. Also, some traits may be a matter of degree. For example, orderliness (organization) is sometimes cited as an indicator of effectiveness, but too much orderliness may result in a student's reliance on the teacher as a substitute for his own self-reliance (Hamachek 1975b, p. 317). Donald Medley (1971) in a review of effectiveness research claimed that at that time none of the following had been found to be related to teacher effectiveness: experience, frequency of work with pupils, knowledge of subject, praise-approval-warmth, indirectness, and pupil participation. Dunkin and Biddle (1974) also found that there were at best weak, and at times negative relationships between pupil achievement and indirectness, praise, acceptance of pupils' ideas, amount of teacher talk, and pupil talk. Moreover, they noted that teacher questions were not found to be associated with pupil attitudes. Flanders and Morine (1975) found no studies which showed that making plans for instruction resulted in improved learning.

Thus it appears that the information dealing with the effectiveness

offers few concrete recommendations for the evaluator, but does bring to the surface evidence which contradicts other findings and which also runs contrary to conventional wisdom. The problems may not be due so much to a general lack of any overriding criteria but rather to a wide variety of interactions involving the teacher, the content, the students, the environments, and the goals of the school (Ritenour 1977, Krasno 1972).

Evaluation Techniques

Having considered evaluation from a very general viewpoint, and also having discussed the material available on teacher effectiveness, it seems appropriate to shift our focus to the procedures which are being and can be used in actual practice in a teacher evaluation program. Several sources of information will be considered, including supervisors, peers, students, and the teacher himself. In addition, several means of gathering data will be reviewed including ratings, observation, video-taping, and self-appraisal.

Administrative ratings

> Administrators continue the semi-annual ritual of writing narrative reports and/or checklist evaluations on teachers. These 'evaluation' devices generally not only fail to measure adequately professional competencies, but also actually result in alienating the relationship between the teacher and the administrator, do little or nothing for improving performance, and engender a false sense of security about the quality of professional performance in the school system (Lewis 1973, p. 11).

By far the most common method of teacher evaluation involves a supervisor, usually a principal, making classroom observations and recording his reactions on some kind of rating form. Such procedures

are not well received by the teachers and have little good to be said for them in the literature, yet they remain. Ratings in general have been cited for their inaccuracy (Powell 1970). They also fail in the process of "weed(ing) out even those most in need of weeding" (Popham 1975b, p. 288). Popham (1975b) has pointed out that ratings have little to recommend them except the convenience with which they can be secured. Other advantages include the possibility for comparisons between evaluators provided that a common form is used, and the relative low cost of administration (Herman 1973).

It is important that those using rating instruments distinguish between professional judgments and their own biases (McNeil 1971b). "In a larger sense, a supervisor who rates a faculty member on his perceived 'goodness' is using institutional self-perpetuation as the ultimate criterion" (Cohen and Brewer 1969, p. 58). It is generally agreed that the person who should do this evaluation is the teacher's supervisor. This person supposedly has the training and job assignment to make him best qualified, and he is the one who is accountable for providing the evaluation. In spite of claims to the contrary, the supervisor is thought to be the least biased, and most authoritative source for teacher evaluation (Herman 1973).

Instrumentation is another important consideration. "When anyone visits a classroom and observes what goes on, whatever his purpose may be, the chances are that he can achieve it better if he uses an objective instrument to guide his observations (Medley 1964, p. 273). Two conditions have been raised in this regard by Bradley et al. (1964). First, factors which can be observed most consistently and with the least "halo

effect" are those low inference items which are most objective. Ratings on personality variables are the least consistent. Also, in any observational scheme it is necessary that the observers receive specific and thorough training in order to obtain reliable, valid, and discriminating results (Bradley et al. 1964, Ryans 1964). Popham (1975b) observed that in reality the practicing educator's knowledge regarding teacher evaluation is extremely weak. Ryans noted: "Much of the ambiguity of the data based upon direct observation and assessment appears to arise from the lack of common understanding and procedure on the part of different observers" (Ryans 1960, p. 72-73). Even though the context of this remark was the use of observation for research, his perceived need for observer training should not go unheeded for those who have the responsibility of evaluating teachers.

> Only with training of observers can one expect to obtain meaningful assessments of teacher behavior. It is the only proper way one can approach teacher assessment for either research purposes or for pre-service and in-service teacher evaluation (Ryans 1960, p. 107).

In spite of its widespread acceptance, criticism abounds for this method of evaluating teachers. One problem "stems from the fact that an extensive amount of important information about what has occurred in a given classroom is usually reduced to a number of highly subjective and impressionistic endorsements on imprecise scales" (Pedersen 1975, p. 16). Another "limiting factor may be the practical fact that the appraiser is just not expert enough to help some highly competent teachers" (Commission of Public School Personnel Policies in Ohio 1972, p. 2). These criticisms could be balanced by having complete job descriptions, well-designed instruments, and well-trained observers.

"The most important criticism of classroom observation for evaluation is to be found in the contrast between the amount of time necessary to secure a valid sample of the total classroom behavior and the amount of time that is usually spent" (Commission of Public School Personnel in Ohio 1972, p. 2). It has been suggested that teachers should be observed in the classroom several times each year at different times of the day and during a variety of types of instruction (National Center for Educational Communication 1971). Unfortunately, it is more common for a teacher to be observed on one occasion for a period of one hour or less during an entire school year.

It has been suggested that "If ratings of effectiveness are valid they should show at least moderate correlation even with imperfect criteria" (Medley and Mitzel 1963, p. 257). Medley and Mitzel (1963) have included quotations from eight "typical" studies involving teacher ratings, all of which show little or no relationship between the ratings and other criteria such as pupil achievement and pupil gains.

Other problems involve the interaction of the personalities of the teacher and the evaluator (Sinatra 1975, Bradley et al. 1964).

> The assumption has seemed to be that if the teacher has a friendly personality and respects the personality of the pupil and that if she is active, enthusiastic and in good standing with the others of the school personnel and in the community, then she is an effective teacher (Worcester 1968, p. 127).

Gage (1972) claims that the presence of an observer who can effect one's standing may test a teacher's nerve to a far greater degree than his skill. "In summary, despite their prevalent use throughout the field of education, ratings have proved almost worthless in isolating teaching competency" (Popham 1971a, p. 9).

Since this is the case, what can be done to improve the situation? Earlier in this section it was noted that job descriptions, improved instruments, and observer training were necessary. Other things which should also be considered are increasing the frequency of observations, avoiding atypical situations, observing a wide range of activities or classes, repeating any questionable observations, observing for varying lengths of time including several whole class periods as well as shorter time periods, and checking observations against the observations of another administrator (Ellman 1976).

Student ratings

> It is well to remember that student evaluation is continuous and inescapable. The only question is whether or not we care to know what it is. E.R. Guthrie (Slobin and Nichols 1969, p. 247).

Student evaluation of teachers offers some interesting advantages which at least on the surface make it most tantalizing. The observation is continuous, a large number of individual biases can be averaged, and there is virtually no dollar cost. Yet student rating of instruction has not as yet received widespread acceptance, especially in the public schools. The following quotation from a study of 286 vocational teachers in high schools or technical institutions in New Jersey indicates that at least some are not pleased with this situation.

> Of interest... is the fact that within the educational milieu, the only source of feedback to teachers, typically, are (sic) their supervisors. The data collected here indicate that such feedback is doing more harm than good, with the 'best' source of feedback, students, overlooked (italics mine) (Tuckman and Oliver 1968, p. 300).

Even critics of student evaluations have indicated that the student is

in the best position to evaluate the worth of the teaching product, being its primary consumer (Rodin 1973, Popham 1975b).

Use of student evaluations appears to be spreading. In 1970 a National Education Association survey of 500 school districts indicated that only five had student evaluation of teachers, and in four of these the practice was voluntary (Poliakoff 1973). A sampling of school boards in 1971 by the National School Board Association of teacher evaluation practices omits, with one exception, any mention of student evaluation of teachers (Poliakoff 1973). "By the middle of 1973 the Educational Research Service found that nearly one out of four school districts surveyed ha(d) some form of student evaluation of teaching" (Halbert 1975, p. 3). Poliakoff has pointed out: "Student evaluation may not have a place in the public school system simply because no one with power has ever asked for it. If that happens, individual school districts do not have many places to turn for guidance" (Poliakoff 1973, p. 42).

Much of the research on student evaluations has been done at the college level and does not generalize to lower levels, especially since there is not much agreement about the value of such evaluations at the college level. Rodin and Rodin (1972) in a study involving teaching assistants of calculus classes found a strong negative relationship between a measure of student learning and a teacher evaluation performed by students. Frey (1973), in a study designed to correct what he considered to be the weaknesses of the Rodin and Rodin study found that ratings of students in a calculus class were correlated with class performance on a common final examination, and that ratings on several

instructional factors were highly related to class performance even though they did appear to be independent of the student's own grades. Several other studies have indicated that there is a positive link between high ratings for instructors and student achievement (Brown 1977, Bryson 1974, Doyle and Whitely 1974). Hildebrand has taken an even stronger position: "I know of no study that shows another method to be as valid and effective as student ratings for incorporating evaluation of teaching into promotion procedures" (Hildebrand 1972, pp. 56-57).

There is evidence that student ratings do reflect more than just differences which can be accounted for in the effectiveness of the instructors. Conditions such as class size, elective versus required, time of day, and other factors have also been reported (Gage 1972, Cornwell 1974). However, in a study done by the Educational Testing Service which involved comparing the responses of current students with alumni who had graduated five years previously, where respondents were asked to name the best and worst teachers they had had, showed high agreement between students and alumni regarding the same teachers (Ritenour 1975).

The question then becomes, what should students be allowed to evaluate? "Students are capable of evaluating much more than we permit. On the whole, they evaluate what we let them evaluate,..." (Dressel 1976, p. 346). "The student probably is the best judge of whether he was bored or stimulated, but is the student the best judge of his ability to comprehend science materials or to apply scientific principles in a new situation" (Rosenshine and McGaw 1973, p. 151)? Since this leaves a relatively

open door, the next question is what do student ratings tell us? A study of 1427 classes including 142,810 students in grades seven through twelve from a six state area responded to Western Michigan University's Teacher Image Questionnaire. Factor analysis revealed that "the single most important factor is viewed here as a kind of teacher charisma or teacher popularity" (Coats and Swierenga 1972, p. 359-360). This factor accounted for more than sixty-one per cent of the total variance! Yet looked at another way, the authors suggested that this left about forty per cent of the variance unaccounted for by this factor, and concluded that student evaluations could still add valuable information when used as part of a total evaluation package.

One problem with applying student ratings to levels below college rests on the argument that younger students are not mature enough to make good raters. In contrast to this, Jones has said: "On the average, secondary students do a more accurate job of rating teachers than do supervisors, other teachers, or principals" (Jones 1972, p. 474). Haak et al. (1972) have done considerable work with students in grades below high school. Their findings are even more surprising:

> It would seem quite sound to presume that their abilities (to discriminate) are entirely adequate - on a group basis - for assessing the quality of teacher-pupil interactions present in a classroom (Haak et al. 1972, p. 11).
>
> At the present time, there is no reason to suspect that the ratings of young students are any less reliable than the ratings of older students (p. 12).
>
> There appears to be no real question of the validity or usefulness of children's perceptions of teachers (p. 13).

If student ratings are to be used, then it is imperative to decide what purpose they should serve. Slobin and Nichols (1969) have suggested two possibilities: to help improve content and methods and to help

administrators in the selection, placement, and utilization of teachers. The former appears to have the most support. Studies cited by Eastridge (1976) indicated that high school teachers improved their ratings as a result of student feedback, and that student feedback was superior to supervisory evaluations when it came to producing positive changes among teachers. Also, a study involving sixth grade teachers indicated that teacher behavior could be changed by giving them a brief summary of information about pupil's opinions. Other sources have also reported that student feedback often yields improved teacher performance (Gage 1972). On the other side of the coin, Vogt and Lasher (1973) in a study at Bowling Green failed to find any significant relationship in teacher ratings over time indicating that there was no improvement after feedback. In support of a program which recommends using student ratings twice during a term, once early and the other at the end, Miller (1972) noted that there tends to be a fairly high correlation between the ratings, indicating that there is little improvement. The strongest statement has been made by Kulik and Kulik: "There is no convincing evidence that teachers use information available on student ratings to improve their courses or their ratings" (Kulik and Kulik 1974, p. 56).

Critics of student ratings of teachers have raised several other arguments which should be considered. Some claim that a "halo effect" based upon popularity of the instructor will color student evaluations (Coats and Swierenga 1972); others would claim that the students are not able to judge the merits of teaching in light of educational goals (Bradley et al. 1964); still others point out that at least at lower levels, a student's positive perceptions can be blotted out by some overriding

negative characteristic of the teacher (Hamachek 1975b); some even claim that the only way to determine whether student evaluations are to be relied upon is to wait until the student can determine the influences of his various instructors have had on his life (Chisholm 1977). To these critics Poliakoff replies: "The disadvantages are obscure. They apparently center around not knowing the value of student ratings or their effect on teacher performance and student-teacher relationships" (Poliakoff 1973, p. 5).

In conclusion,

> if the evaluation is designed as a personality questionnaire, a fault-finding expedition, or a form of retaliation, then the validity of the student evaluation must be seriously questioned. If the evaluation is designed for the improvement of instruction, then the process has the potential for making positive contributions to the growth of both students and faculty (Halbert 1975, p. 9).

Systematic observation

> The fact that ratings of teacher effectiveness have no discernable relationship to effectiveness does not mean that effectiveness cannot be measured in process. Since it may be assumed that whatever effect a teacher has on pupils must result from his behaviors, it is only necessary to identify the crucial behaviors, record them, and score them properly to measure effectiveness in process (Medley and Metzels 1963, p. 258).

In an attempt to improve the reliability and validity of teacher rating systems, several researchers have designed and applied various systems for the purpose of "scientifically" measuring teacher behavior. Hayman and Napier have stated:

> Observational systems attempt to isolate sets of mutually exclusive behaviors, which are organized so that a trained observer is able to understand how a person (or group) is behaving during a certain period of time; the systems describe

> what occurs during a particular time period as accurately as possible. By categorizing what is to be observed, the observer becomes a detective looking to discover the behavioral clues that will lead to an understanding of what happens in the teaching-learning situation is, in fact, occurring (Hayman and Napier 1975, p. 91).

Most classroom observation systems, whether they are effective, cognitive, or multidimensional, require an observer who employs a systematic method of recording teacher and student behavior. Most of the systems record only verbal behavior (Sandefur and Bressler 1971). The first study of any magnitude to determine patterns which discriminate effective and ineffective teachers was done in 1929 by Barr (Medley and Mitzel 1963). In 1945 a study by C.D. Jayne revealed that "individual items which themselves do not differentiate between teachers or classes can often be combined into sets of items or scales which do" (Medley and Mitzel 1963, p. 261). Early developmental work with affective systems was done by H.H. Anderson and John Withall. The system developed by Ned Flanders has become best known. Other affective systems which constitute expansions of Flanders have been done by Amidon and Hunter. Cognitive observation systems have been developed by Bellack, Smith and Meux. Multidimensional systems have been developed by Spaulding, Medley and Mitzel, Ryans, and Openshaw and Cyphert (Sandefur and Bressler 1971). Simon and Boyer (1967) have listed and described seventy-nine different observational systems that are in existence. Not all of these were intended for evaluation purposes, but many could be adapted.

In constructing an observation system to be used in teacher evaluation, it is first necessary to classify teaching behavior into a number of response categories. "Once categories of teaching behavior have been chosen, it is then possible to construct instruments by which to assess

the behavior of the teacher in these categories and to measure the effects of such behavior on students" (Meux and Smith 1964, p. 163). The Flanders system is not intended to be a measure of successful or unsuccessful teaching, "rather it is designed to be an objective indicator of the type of verbal interaction going on in the classroom" (Commission of Public School Personnel Policies in Ohio 1972, p. 8). However, Popham has noted that "the architect of the observation scheme clearly is obliged to make judgments regarding what kinds of teacher behaviors contributes to effective or ineffective instruction" (Popham 1975b, p. 288).

Flanders and Morine have found at least one valuable use of systematic observation:

> One general conclusion from all the research which uses some form of interaction analysis is that when teachers or college students take the time to analyze their own patterns of verbal interaction, they are likely to change these patterns.... This generalization is supported, to one degree or another, by no less than twelve different research projects (Flanders and Morine 1975, p. 87).

Marten (1975) also reported that teachers who used systematic observation techniques have significantly more favorable and stronger attitudes about classroom observation leading to improvement than those who do not use a systematic observation system.

Systematic observation is not without its critics. Rosenshine and McGaw have stated: "At present, most of the advocacy for transaction accountability appears to rest on unverified but implicit assumptions that specified educational practices will lead to outcomes of worth" (Rosenshine and McGaw 1973, p. 149). Powell (1970) has even challenged the idea that interaction is one of the defining characteristics of

teaching. Another major concern with this approach is its process-orientation, when there is little evidence showing high and consistent connections between what is observed and measures of student outcomes (Popham 1971b, Rosenshine and McGaw 1973, Popham 1971a, Worcester 1968, Popham 1973c). Another criticism is that the process focuses on certain behaviors to the exclusion of other negative behaviors which may outweigh the effect of those which were recorded (Popham 1975b). Finally, "the standard of excellence in teaching commonly held (flexibility) implies a teacher whose behavior is inherently unstable. Needless to say, this poses a problem for an observer trying to observe a teacher's customary and usual ways of teaching" (Berliner 1975, p. 11). "The danger in using these systems is that one begins to believe that the presence of a statement in certain of the categories called for is automatically good or bad" (McNeil 1971b, p. 65).

Other process evaluation methods

Several other sources of evaluative information exist. One sometimes suggested is the use of peers. Cohen and Brewer believe that peer evaluation "is the scheme least likely to meet with resistance" (Cohen and Brewer 1969, p. 10). Advantages of peer evaluation center around the fact that a person in the same general academic area should possess more in-depth knowledge of the subject, and that he is best able to assist a colleague in a nonthreatening fashion. On the other hand, peers lack the authority and responsibility to evaluate, they may not be objective, their ratings may conflict with those of the supervisor, and peer evaluation has even been thought to lead to inter-worker conflict (Herman 1973).

Elliott (1974) has recommended that a panel of teachers selected by the faculty serve as a pool from which a principal would choose evaluators. This would allow for multiple evaluators and multiple visits, with a net result being a plan of growth and development for each teacher. However, "the feedback that peers offer is conceivably valuable but, like the 'round robin' exchanges that typify certain adolescent searches for self-knowledge, it is likely to be fraught with subjective, nondirective assessments" (Cohen and Brewer 1969, p. 10).

Another possible method of evaluation involves asking qualified observers for reports of incidents which were 1) negatively effective, leading to failure, and 2) positively effective, leading to unusual success. This looking at "critical incidents" allows the evaluator to notice frequent behaviors in either category and recommend constructive change where needed (Remmers 1963).

Process evaluation: a summary

> There appears to be a consensus that the focal point for evaluation be what the teacher does in the classroom in the performance of instruction. However, there is little agreement on the criteria to be used for judging what should be done in the classroom (Bolen 1973b, p. 72).

Measuring teacher performance rather than pupil behavior rests on the assumption that there exists a basis, a reliable relationship between assessed teacher behavior and pupil behavior. At this time, no validation exists (Scott and Thorne 1974). "When measuring process we have no guarantee that we are measuring what we want to measure (i.e. effectiveness of instruction)" (Weisenstein 1976, p. 13). However, "for the improvement of instruction, process evaluation is far superior to product

evaluation" (Medley 1973, p. 33). Carrying this to the extreme, Powell has said, suppose

> a teacher does all that anyone could possibly do and yet fails to bring about learning, or even to evoke any response from his pupils. Such a teacher would certainly be judged unsuccessful but it would be grossly unjust (and absurd) for him to be judged to be incompetent (Powell 1970, p. 138).

One suspects that there might be room for argument here.

Product measures

> ..., if teachers must be judged... let them be evaluated on the effects of their efforts, not on perceived worth of the efforts themselves (Cohen and Brewer 1969, p. 65).

> For many years researchers have sought to identify the characteristics of the effective teacher; more recently, attention has turned to analysis of teaching behavior. None of these efforts should obscure the fact that pupil learning and behavior are the purpose of school and, therefore, must be the ultimate objects of evaluation (Howsam 1973, p. 14).

This reference to product measures as the "ultimate criterion" for evaluating teacher effectiveness has been raised by many authors (Hildebrand 1972, McDonald 1972, Bolton 1973b, Lauritz 1967, Herman 1973, Saadeh 1972, Cohen and Brewer 1969). Weisenstein has said:

> It would then appear that since the act of 'good' teaching is such an illusive concept, evaluation of the student product would yield more valid data regarding instructional ability and would be more readily interpretable to lay persons (Weisenstein 1976, p. 15).

California's Stull Act of 1971 caused more than a little distress when it required the use of pupil progress toward district standards of achievement in all areas of study be a part of teacher evaluations (Popham 1973a).

To even strengthen the contrast between proponents of process

evaluation and those favoring product measures, Cohen and Brewer have stated: "... effectiveness should be measured only in terms of what eventually happens to the end products - the dependent variables, the students' learning. Faculty performance may or may not be relevant" (Cohen and Brewer 1969, p. 4). The essence of the argument lies in the position that any evaluation, whether of students or teachers, can only be justified in terms of learning (McKeachie 1970).

There are some clear advantages to this approach. First of all, it places emphasis on the results of teaching rather than the intentions. Second, it focuses on an examination of the needs of the pupils. Third, pupils are more involved, and this allows for an even closer review of teacher performance (Carlson and Park 1976).

In spite of the apparent agreement on many fronts regarding the use of measures of student growth as an important criterion in evaluating teacher effectiveness, a study by Miller and Miller (1971) involving administrators reacting to a questionnaire dealing with personal qualities and professional competence of successful teachers revealed that class achievement was ranked last by elementary principals and in the lower one-third by secondary principals and superintendents when ranking professional competencies. Leading the list was classroom management and discipline and knowledge of subject matter.

Some research has been reported that supports the use of pupil achievement in teacher evaluation. McNeil reported:

> The data in the studies... provides evidence that the emphasis and use of operational definitions of instructional goals, including specification of criterion measures, in the supervisory process is accompanied by more favorable assessment of teachers by supervisors and greater gain in desired directions on the part of learners (McNeil 1967, p. 71).

A study performed by Moffitt and cited by Lucio (1973) indicated that teachers evaluated by agreed upon instructional objectives 1) had pupils who performed better on a post-test, 2) had more confidence and satisfaction with their supervisor, and 3) preferred performance evaluation based on pupil achievement as opposed to rating scale measures.

Several options have been suggested for obtaining product measures. These include standardized achievement tests (norm-referenced), the use of student gain scores on norm-referenced tests, and criterion referenced instruments (Neel 1972).

Norm-referenced tests themselves have fallen into disfavor among a growing number of educators for reasons which go beyond teacher evaluation. One common complaint is that norm-referenced tests are more a measure of I.Q. than of achievement, especially achievement attributable to a teacher (Popham 1975b). For the purposes of teacher evaluation "this method of appraisal can be used with justification only when the achievement of students under various teachers can be objectively appraised and corrected for factors beyond the teachers' control" (Gage 1972, p. 172). For one thing, pupil achievement is due at least in part to the instructional environment provided by previous teachers (Herman 1973). Other out-of-school variables also enter into the picture, but "the present rudimentary state of our quantitative models does not permit us to disentangle the effects of home, school, and peers on students' achievement" (Mood 1970, p. 7).

Another problem arises because of a conflict of values: different teachers often seek to accomplish different objectives (Popham and Baker 1966, Pedersen 1975). Berliner has noted that standardized tests "may

not reflect what was taught in any one teacher's classroom.... They simply lack content validity at the classroom level" (Berliner 1975, p. 4). As a result, in order to have a favorable evaluation, there might be a tendency to "teach to the test" (Soar and Soar 1975, Herman 1973, Medley et al. 1975). This tends to make the focus of instruction too narrow, and cause other objectives, which may be longer ranged but harder to measure, to be overlooked.

> To evaluate teachers and schools solely on the basis of the subject matter gains made by pupils grossly under-represents the broad range of objectives for which teachers and schools have been given some degree of responsibility. Yet for many of these objectives there are no measures which are immediately, for some even remotely, available (Soar and Soar 1975, p. 17).

Medley adds:

> ...if we are measuring teacher effectiveness for evaluation purposes, as I assume most of you will be, we need to measure effectiveness in achieving most, or at least a good share, of the things teachers are supposed to do. If we include ability to help pupils develop attitudes and values or acquire inquiry skills (for example) as a part of what an effective teacher does, it is quite clear that measures of pupil gains are and must for a long time remain lacking in content validity because of the lack of valid tests of these characteristics (Medley 1971, p. 10).

Krasno has pointed out that focusing on immediate effects of schooling neglects the long term consequences, but that "many of the most profound objectives of education are expressed in terms of the life-long impact of schooling" (Krasno 1972, p. 3).

One attempt to improve the use of achievement tests as measures of pupil growth has been to use a pre- and post-test approach. McKeachie has even suggested that this could be lengthened beyond one year by testing students' "interests, skills, and knowledge before the course

and at several points in time after the course" (McKeachie 1970, p. 2). This allows for an analysis of teacher effects which use gains in achievement levels, not just achievement levels.

Problems arise even with this approach. First of all, the tests themselves may not adequately measure learning. Second, classes are not randomly assigned, giving some teachers an unfair advantage (Glass 1974, Burnett 1975). Medley (1975) has noted that reliability of test-retest using standardized tests has been reported to be only about .3, thus ninety per cent of the variance is due to other sources, leaving only ten per cent accounted for by teacher competence. Roth and Mahoney (1975) have reported that post-tests are more related to pre-tests than to any measure of teaching, showing that the gain is more often a function of student ability. There are also problems of regression to the mean (Roth and Mahoney 1975) and instability (Medley 1971). Another problem which has deep philosophical underpinnings is the fact that measuring teacher competence through pupil gain in high level objectives is difficult and may be impossible due to the lack of measuring instruments available (Roth and Mahoney 1975).

One attempt to improve on the above has been suggested by Burnett (1975). This author's approach is to use class averages of pre- and post-test results for all teachers in a district to plot a regression line. Teachers above the line would then be better than average and those below would be poorer than average. Unfortunately, no matter how good the teachers in the district might be, this scheme guarantees that about half must show up below average!

A third method for testing student learning is criterion-referenced

(or domain-referenced) testing. This type of testing is based on the theory that "perhaps the most important skill of a teacher is the ability to bring about changes in the behavior of students on prespecified objectives" (Millman 1974, p. 392). "The really central attribute of a domain-referenced measure is that it is based on a detailed description of the nature of learner behavior to be assessed" (Popham 1974b, p. 37). Thus criterion-referenced testing attempts to determine the level of accomplishment of the learner against some established standard, while norm-referenced testing attempts to discriminate between learners. The problems of criterion-refererenced testing focus on the types of behavior which is tested.

> ...(I)f you stay with behavioral objectives that deal, as most of them do, with cognitive aspects (that is, knowledge of content, general ability in processing data), if you stick with these alone, you miss a number of important educational outcomes (Hastings 1973, p. 142).

Above and beyond the measurement questions that have been raised here, other considerations might be made, such as how a teacher performs his out-of-class duties (Bradley et al. 1964). However, Bohlken and Giffin (1970) have concluded that little negative criticism can be raised in regard to student growth criteria <u>if</u> adequate instruments for the measurement of the prescribed educational objectives are available and <u>if</u> casual behavior on the part of the teacher can be determined. There is little in the literature to remove these two conditions at this time. Medley has flatly stated: "On the whole, I think we should give up the idea of measuring teacher effectiveness in terms of pupil gains on tests, attractive though the idea may seem on the surface" (Medley 1971, p. 12).

Performance tests

> Admittedly, pupil growth is the ultimate criterion for assessing teacher effectiveness. However, it is unsound to rank teachers on this criterion when they have not been confronted with a comparable set of teaching conditions including factors such as common instructional tasks, teachable children, and time allowed for teaching. The problem, therefore, is to design tests of teaching power by which teachers have an equal chance to show their relative ability (McNeil 1971a, p. 1).

One of the major complaints against most of the evaluation procedures discussed above is that they are attempting to make comparisons in situations which are by their very nature diverse. "What is needed is a procedure for simulating the problems a teacher encounters when he interacts with a class, a procedure which can be duplicated over and over so that more than one teacher can be confronted with the identical problem" (Roth and Mahoney 1975, p. 29). One solution to this request is the teaching performance test. This test gives a teacher one or more explicit objectives plus a sample test item. The teacher then prepares a brief lesson for teaching the objective(s) (background information is supplied to the teacher as needed). The teacher then gives the lesson to a class which is chosen with careful statistical controls to insure a representative sampling of students. A post-test is administered to the class to determine the amount learned, and the instructor is also rated by the students based upon their interest level (Popham 1975a). "The problem of different objectives is hopefully alloyed by giving teachers identical goals to achieve" (Popham and Baker 1966).

Teacher performance tests can be contrasted with the more commonly used micro teaching in that teaching performance tests focus on the product of instruction while the micro teaching focuses on instructional

acts (Bolen 1973b). Micro teaching is generally scaled down teaching, which usually involves normal subject matter and possibly statistical control to allow for comparisons across teachers (Gage 1972). Both micro teaching and performance tests can be used for inservice purposes to expose teachers to a wide variety of situations while providing for control (McDonald 1973, Popham 1975a).

Teaching performance tests have been used by several authors including McNeil (1971a) and Popham (1971a). A report of the use of such tests in a research project can be found in O'Connor and Justiz (1970), and Popham (1971b).

Even though the proponents of teaching performance tests have expressed great hopes for these devices, the instruments have come under fire from critics. Even Popham (1975b) has admitted that it is too early to tell if the tests measure well enough to distinguish between teachers. Glass also reported that "the technique has not been shown to possess reliability adequate for measuring individual differences among teachers" (Glass 1974, p. 16). Glass went on to say that performance testing lacked information across different topics and across different groups of students. One of the problems lies in the admitted difficulty of the developers to find topics which are suitable for the tests (Popham 1975a). In the same article, Popham admitted that there were also logistic problems, especially if there was to be reteaching.

Other critics have questioned more far reaching effects of this approach to evaluation. Soar and Soar have written that "there are questions of whether teaching material which does not have to be integrated into previous knowledge requires the same skills as the usual teaching

setting and whether such short-term learning generalizes to long-term learning" (Soar and Soar 1975, p. 10). Berliner (1975) has raised similar doubts, bringing into question the predictive validity of the tests. In defense, Popham has countered:

> It should be emphasized, however, that the ability to accomplish pre-specified objectives in learners under short instructional periods represents only one criterion by which a teacher should be judged. The use of teacher performance tests simply reflects an additional criterion which might be employed in a local district's evaluation system (Popham 1971a, p. 39).

Self-evaluation

> Effective instructors are constantly attempting to improve themselves. An active process of continuous learning to use himself as a more effective tool describes the good teacher's reaction to what evaluation means to him (Hanke 1973, p. 56).

Authors have noted that the best judge of teaching are teachers themselves (Vail 1974), and that teachers must make the changes if they are to make improvements (Crim 1974). Flanders has written:

> The most effective changes in methods of instruction occur when a teacher can compare what he wanted to accomplish with a nonthreatening, objective summarization of his spontaneous behavior. Using proper procedures, the teacher can make his own discoveries and reach his own conclusions about what changes would reduce any discrepancy between intent and action (Flanders 1964, p. 224-225).

This position is accepted by Horton (1977) and McNeil (1971b). In fact, Horton has stated that "change occurs only when the individual recognizes the need for change..." (italics mine) (Horton 1977, p. 6).

> Self-assessment is probably the most powerful means yet developed for a teacher to be the master of his own professional growth. Self-assessment is bold but easy to understand, revealing and thus threatening, majestic in goal and thus giving dignity to the teaching profession (Bodine 1973, p. 171).

The appeal here and elsewhere (Bolton 1973b) is for a means of evaluation which does not expose the teacher to an outside threat.

Bodine (1973) has suggested five steps to be followed in the self-assessment process: 1) the teacher learns to use measurement instruments; 2) these measures are applied to teaching performance; 3) the teacher states one or more goals in measurable terms; 4) the teacher tries out both the instrument and performance of the goal(s); and 5) the process (steps 3 and 4) is repeated. Bodine contends that such a practice gives the teacher accurate feedback of existing behavior while producing an evaluation which is made in an emotionally supportive atmosphere. The anxiety produced is a result of comparing what is to what is hoped, and is thus free from outside threat. Redfern (1972) has elaborated on this:

> Self-evaluation, properly used, is a guide for planning further self-improvement. It is not a device for self-incrimination, providing damaging evidence which might be used by the principal or his superiors to injure the teacher's professional status in some manner (Redfern 1972, p. 42).

Self-evaluation is not without its problems. Remmers (1963) and Herman (1973) have mentioned the tendency of a person to overrate himself. Crim has seen another set of difficulties:

> There are two basic problems involved in self-appraisal: (1) the inability of the teacher to see and hear himself as others do, except vicariously through the reports of others, and (2) the inability to recapture except through verbal vicarious recall what the teacher actually did and said and looked like when it actually happened (Crim 1974, p. 6).

As one solution to this pair of problems, Popham (1971b) has suggested the use of teacher performance tests as described above. Another method, the use of video-tape recordings, will be discussed in the next section.

Video-tape in teacher evaluation

> The use of video-tape as a diagnostic record provides a new dimension for the teacher. As the teacher observes himself on the video-tape and discusses these observations with an understanding principal or curriculum specialist, the present mode of teaching in comparison to the desired mode can be planned. In addition, after a specified period of time, a second video-tape can be utilized to point out the degree of improvement in the instructional process (Wilson 1975, p. 5).

Until fairly recently, it was only possible to make audio recordings of classroom activities. The use of a video-tape recorder (VTR) has added a new tool to the bag of the educational researcher. The uses of a VTR are many. A teacher may receive immediate feedback regarding a particular teaching performance and thus may practice a given skill, or review different strategies used by himself or colleagues (Crim 1974). The VTR also eliminates the problems inherent in human recall while providing an objective record of behavior which is free from the interpretative judgment of another person (Crim 1974). By using two cameras and a split screen, it is even possible to record both teacher behavior and student response (Baltus 1974).

Video-taping also fits well into several evaluation modes which have already been discussed. The use of systematic observation instruments allow the teacher to observe more than just methodology, but rather to view verbal and nonverbal behavior and student reaction. Coding then gives the teacher a common language for discussing the classroom interaction with others (Crim 1974). Video-tape also helps the teacher see personal mannerism and appearances, and amount of teacher talk, and the use made of the blackboard and materials (Herman 1973). The use of the VTR has been encouraged in Nebraska where the Video Inservice Program

of Nebraska Educational Service Unit 6 has a team which trains school staff members to operate video equipment for the purpose of self-appraisal, including the use of self-appraisal instruments, identifying job target behaviors, and general self-appraisal counselling (Poliakoff 1973). Ellett and Smith (1975) have indicated that inservice training can be greatly enhanced by using video-tape in self-assessment, but that this will require a continuing effort on the part of supervisors to provide the support systems and encourage their use. Mallery has even suggested that video-taping is "a way of coming at student evaluation of teachers that focuses on how we - students and teachers in a given class - are doing, rather than a sudden, unprepared analysis by students of how they think the teacher is doing" (Mallery 1975, p. 28).

The use of the VTR has some disadvantages. A major one is that people do not react normally in the presence of the equipment (Herman 1973, Commission of Public School Peronnel Policies in Ohio 1972). Also, the equipment is costly and operators must receive adequate training (Baltus 1974). The major advantage seems to be in seeing oneself from the viewpoint of the students. Smith has noted that a person

> observing himself on a video-tape recorder generally increases his perceptions of what he thought he was doing which may be different than what he actually did. He also discovers, when viewing himself on a VTR, that his perceptions of himself differ from his pupils' observations of the same behavior (Smith 1974, p. 30).

Performance evaluation

> It's time we focus, together, on an evaluation scheme such as Redfern's, which begins with goal setting, continues with strategies for achievement, analyzes whether achievement was reached, and begins again with review of goals and strategies

in another cycle.

> That is the framework which individual states and school districts might use in the development of their own system of teacher evaluation (Pierce and Smith 1974, p. 6).

To complete the study of the various methods of teacher evaluation, we shall look at performance evaluation, often called the Redfern Approach, Management by Objectives (MBO), Supervision by Objectives (SBO), or performance appraisal. Gray has defined performance appraisal as follows:

> ...a positive, systematic, individualized due process evaluative program that can be applied to all members of the school organization. It is based on the assumption that people want to do a good job. It puts responsibility on both appraiser and appraisee to reach mutually agreed upon objectives (Gray 1975, p. 3).

The thrust of MBO is to channel all efforts of the school system toward achieving specific results in an established time (Lewis 1973).

Several models have been suggested in the literature. Perhaps the one most often cited is that of George Redfern. The Redfern Model consists of a six step process. The first component involves establishing performance criteria which are the duties and responsibilities required in the performance of the job specified. The second component involves setting specific performance objectives or Job Targets. The third component is determining the performance activities which are designed to attain the objectives. The fourth component is the monitoring process which includes the collection of data relating to the attainment of the performance objectives. Component five is the assessment of the data, and is the key to the scheme. Finally, component six is a conference where constructive criticism is given and a follow-up program is determined (Redfern 1972). Other similar procedures have been suggested by Saif (1976), Spivey (1976) and Thomas (1974).

Coney (1972) has listed six rules for performance evaluation which clearly distinguish it from the other types of teacher evaluation. First of all, in performance evaluation the job is evaluated, not the person doing the job or even the methods used. Second, focus is on desired results. Third, the use of cooperative goal setting provides a means of assuring that there will be agreement on what the objectives are. Fourth, the standards themselves specify what is to be done and also the means for determining satisfactory performance. Fifth, performance evaluation assumes responsibility for providing those whose performance is evaluated reasonable assistance to help them achieve success. Finally, performance evaluation gives each member of the team an opportunity to improve, which is the purpose of most evaluations (Coney 1972).

There are some very important assumptions which underlie the MBO approach. "It does require a climate of trust and a professional attitude on the part of administrators and staff that instructional improvement is the real, and not the imagined goal of an evaluative procedure" (Spivey 1976, p. 44). "There is also an assumption on the job targets approach that performance appraisal is continuous throughout a person's career - that tenure does not imply that an administrator or teacher need not define job targets and strive to meet them" (Poliakoff 1973 a, p. 1).

The performance appraisal model requires that administrators become involved with teachers in a partnership for the purpose of evaluation (Poliakoff 1973a). The supervisor must also possess good human relations skills, since those evaluated must be involved in selecting their objectives (Hayman and Napier 1975). The focus of the process is establishing instructional goals, defining these goals in terms of program

and instructional objectives, and developing job descriptions in terms of how to accomplish the desired ends (Sarthory 1973).

> The objectives which the teacher sets for himself become part of the controls against which the teacher appraises his performance. These objectives also become the standards and the success criteria against which the teacher's performance will be evaluated during the post appraisal conference (Lewis 1973, p. 31).

McNeil (1971b) has expanded on this with the following comment:

> ... supervision by objectives is a process by which a supervisor and a teacher agree in advance on what they will accept as evidence that the teacher has or has not been successful in changing the skills, competencies, or attitudes of his students. The agreement is drawn up before the teacher acts and is designed to counter the prevailing practice of trying to make ex post facto judgments of ends. The contract is tentative to the extent that at any time the parties can renegotiate (McNeil 1971b, p. 36).

Levinson has also recommended that every performance appraisal program "should include group goal setting, group definition of both individual and group tasks, group appraisal of its accomplishments, (and) group appraisal of each individual member's contribution..." (Levinson 1970, p. 131) whenever more than one person is involved in a particular situation. This is more applicable to a business setting, but can be used in education when there are department or unit goals.

> The philosophy behind objectives is that unless specific objectives on all levels of operating the school system are set, mutually agreed to and performed, there will be relatively little value or basis for measuring the performance of educators (italics mine) (Lewis 1973, p. 31).

Spivey (1976) has listed three sources for objectives: the teacher's own ideas and perceptions of his and his students' needs; the district's, department's, or course description's objectives; and recommendations from the evaluator. These objectives are of two general kinds, the

first being those which deal with aspects of teaching performance and the second involving personal skills which would permit tasks to be performed more efficiently (Lewis 1973). Gray emphasized that "objectives must be seen as relating to on-the-job performance" (Gray 1975, p. 6).

Performance objectives are often broken down into three component parts: a statement of the conditions or constraints under which the activity will take place, a statement of the act which is a specific observable performance, and minimal standards which describe acceptable performance (Lewis 1973, Armstrong 1973). Many of these objectives are product oriented, assuming that learning is evidenced by changes in the behavior of students and that teaching is meaningful only when the teacher's predetermined and intentional changes for the learner actually take place (McNeil 1971b).

The actual performance appraisal process as described above has three key elements. The pre-observation conference is the first step in the evaluation. At this time the teacher and the supervisor agree on what the teacher intends to achieve and this is then written in terms of performance objectives. The participants also agree upon data collection procedures so that the evidence will match with the objectives. Finally, agreement is reached on the role the supervisor is to play during the observations. Gray has written:

> Not only must the appraiser help the appraisee set meaningful objectives, but he and the District must provide the help to enable the appraisee reach that objective. Any other approach would be immoral. Programs that ask you in September to write two or three objectives, come back to a conference in June and tell you how well you did, are doomed to failure (Gray 1975, p. 6).

The gathering of data and making of judgments is the next process. Hopefully the objectives have clearly indicated what is to be evaluated and by what standards. At this point, data can be collected. This information should contain a record of what has occurred, with generalization and analysis to take place at a later time. "Observations should be more descriptive than interpretative, providing descriptive facts with which the teacher can make interpretations and decisions regarding future practice" (McNeil 1971b, p. 60). Data can be gathered using several sources including classroom observation, video-taping, examining lesson plans and materials, and studying examples of students' work (Spivey 1976). Participation by students and peers, and the use of self-evaluation can all be encouraged. The emphasis is on the accumulation of facts upon which decisions can be based.

> ... (T)he judging and analyzing of lessons should follow observations and not necessarily be done simultaneously with observation. Objectivity in observation comes with the recording of concrete, specific behaviors, and the more complete the record, the more objective it will be (McNeil 1971b, p. 64).

Redfern has emphasized that "evaluative estimates should be supportable by evidence of observations made, data collected, conferences held, and assistance provided, all within a framework of fairness and objectivity" (Redfern 1972, p. 43).

The last part of the evaluation cycle is the post-observation conference. At this conference the supervisor and the teacher discuss the degree of attainment of the goals. "Much of the content of evaluation conferences should not be described in terms of problems. On the contrary, the conference more properly is a place where progress is dis-

cussed and understanding is sought" (Redfern 1972, p. 51). To support this, the findings of a study done at a General Electric plant in the mid-1960's revealed that criticism has a negative effect on goal achievement and that the defensiveness resulting from a critical appraisal actually produces inferior performance (Meyer, Kay and French 1964, p. 124). (The same study revealed that mutual goal setting improves performance, and that assistance in the form of coaching needs to be on a day-to-day basis, not annually, to be effective.) McNeil has written:

> Evidence that a good post-observational conference has occurred is seen when the teacher leaves the conference with new objectives he wants to try to achieve, new instructional procedures to try out, and plans for checking the results that follow implementation of the new departures (McNeil 1971b, p. 75).

At this point, the cycle begins again.

There is evidence that this approach to evaluation has some very positive effects, many of them attitudinal. Sources have indicated improved performance (Meyer, Kay and French 1965) and a general improvement in the overall confidence in the appraisal system (Fox and Jones 1970, Eads 1974). This type of evaluation approach has been used successfully in a variety of situations (Watman 1972, Poliakoff 1973, Burnett 1975, Place 1974). "The job targets approach was adopted by 25 per cent of the school systems who reported administrative evaluation procedures in a 1971 survey by the National Education Association" (italics mine) (Poliakoff 1973a, p. 1).

There can be no doubt that this system also has problems. Most criticism focuses on time required, inequities between staff who are pursuing different goals, and the unrealistic nature of some goals

(Carlson and Park 1976). Redfern himself dealt with this when he wrote:

> It cannot be over-emphasized that this kind of evaluation takes time and effort. There are some who take the position that it requires too much of these qualities (sic). There is reason to believe, however, that when the evaluation plan is well-designed and oriented toward the assessment of performance results, it will pay good dividends. Superficial evaluation based upon incidental contacts and cursory judgments may take less time, but it is doubtful that it can or does accomplish as much as that based upon performance objectives (Redfern 1972, p. 27-28).

Setting standards that are too low and thus allowing mediocrity to pass for satisfactory performance has also been cited as a problem (Lewis 1973). Gray has noted:

> In analyzing the targets over a period of years, it was apparent that, as people gained trust in the process, they set more realistic targets. The appraisee, through his own self-assessment, more readily evaluated his own strengths and weaknesses and identified critical areas in which to work (Gray 1975, p. 5).

In conclusion, note the following statement by Redfern:

> The significant point is that a good evaluation process requires a team approach. They (teachers and supervisors) must work together in a mutual effort to improve the work of the teacher in raising the quality of educational performance. This understanding and working relationship is one of the distinguishing marks between performance evaluation and the more superficial rating of teachers (Redfern 1972, p. 34).

Evaluation follow-up

If teacher evaluation is predicted on the need for the improvement of instruction, then making provision for this improvement is a necessary component of any evaluation program. This may well require administrators to reorganize the use of space, time, and resources so that improvement is encouraged. This might also require them to update their own competence, contact colleges regarding new development in instructional procedures,

and provide for outside sources to aid in improvement (Flanders and Morine 1975).

> Since performance evaluation is aimed at improving the program in the school, it is essentially a change process. Inservice education ... can be viewed as an integral part of the change process, providing a method by which staff members become more able to accomplish their job targets (Fast 1974, p. 38).

Herman stated that evaluation must include assessment followed by in-service programs designed to improve those who have been shown to be unacceptable. "...(I)t is grossly unfair to the employee being evaluated if areas of weaknesses are identified and no program of assistance is provided which will enable the employee to overcome his weaknesses and improve his performance" (Herman 1973, p. 194). Weisenstein (1976) has claimed that teachers will work hard at self-improvement if they believe that an inservice program is designed to help them and their students. "Unless the findings from evaluating performances are used to identify and implement indicated programs for professional improvement, the whole process will not have been worthwhile" (McKenna 1973a, p. 23).

Conclusion

Popham (1975b) has given what is still probably the best advice when the body of literature is taken as a whole. He recommended that there be continuous study of the more promising methods of teacher evaluation, that evaluators recognize the deficiencies in the more traditional indices, and that evaluators use multiple evaluative criteria and a variety of techniques.

CHAPTER III. PROCEDURES

In order to determine whether there is any formal evaluation carried on in the Iowa public schools, twenty-five of the largest school districts in the state (based upon total district population) were selected, and each district was contacted by telephone. The district office personnel were asked to respond to the following questions:

1. Do you have written criteria for the evaluation of teachers in your system?
2. Is the same set of criteria applied to all teachers?
3. Has the Board adopted the criteria?
4. Would you please send up a copy of the criteria?
5. Do you have a written evaluation procedure?
6. Has the procedure been negotiated with the teaching staff?
7. Is the procedure part of the master agreement?
8. Is the same procedure applied to all teachers at all levels?
9. Would you please send us a copy of your procedures?

As a result of these telephone conversations, twenty-one of the districts sent materials. These ranged from book-length works to single page evaluation forms. From the response to the questions and after reading the materials received, it was determined that all the districts had some form of formal evaluation. So a study was designed to determine what practices were being followed.

The materials from the twenty-one districts were studied in an attempt to determine the district's agreement with the criteria for a good evaluation suggested by McNally (1973). Based solely upon the

materials provided, points were awarded in each of the following areas:

1. Purposes clearly stated in writing - well-known to evaluators and evaluatees.
2. Policies and procedures reflect knowledge of research.
3. Teachers know and understand criteria.
4. Evaluation cooperatively planned, carried out, and evaluated by teachers, supervisors, and administrators.
5. Evaluation as valid and reliable as possible:
   a. important to learner.
   b. adequate sample.
   c. evaluator agreement.
   d. guidelines or training for evaluator.
   e. limits on range of criteria per visit.
6. Evaluation more diagnostic than judgmental.
7. Self-evaluation important.
8. Self-image of teacher enhanced.
9. Creativity encouraged.
10. Clear, personalized, constructive feedback provided.
11. Part of the leadership role of the principal: part of in-service program.

When the materials clearly indicated that a category was satisfied, a point was awarded. When it appeared that a category might be satisfied, or when a category was partly satisfied, one-half point was awarded. As a result, a scale of agreement with the criteria was produced which had a range of zero to fifteen possible points (see Appendix A).

Once the materials were evaluated against the McNally guidelines,

the school districts were ranked from lowest to highest based upon their scores. The scores ranged from zero for a district whose materials were so limited as to reveal virtually nothing of its program, to a high of twelve and a half. Nine districts were selected from the list so that a range of agreement with the McNally criteria was represented. Each district was then contacted by telephone to ask for their permission to carry out the study. One district declined to be included and was replaced with another which had virtually the same score on the criteria.

A questionnaire was constructed for the purpose of gathering information about each district's evaluation program. A list of 155 commonly used criteria was produced from the materials sent by the twenty-one districts from other evaluation literature. In addition, a variety of questions were written which asked about evaluation procedures, purposes, and staff satisfaction (see Appendix B). The instrument was informally field tested by releasing it to the advisory committee of the author, as well as to a nonrandom sample of graduate students and mathematics faculty at Iowa State University. The instrument was then corrected and duplicated.

In order to have a manageable sample, and yet not to burden any one district, it was decided that each school system would be asked to allow thirty teachers and six building level administrators to participate. In order to make the selections, eight of the districts supplied staff directories. In these cases, the classroom teachers and building administrators were numbered, carefully excluding special education personnel, guidance personnel, and those who teach specialized subjects at the elementary level (e.g. art, physical education, etc.). A random

numbers table was consulted to determine which teachers and administrators would be selected. The sample of thirty teachers and six building level administrators was restricted only in the sense that half of the teachers and half of the administrators were elementary school personnel (defined to be grades kindergarten through six), with the rest secondary personnel (grades seven through twelve). The ninth school district chose to make its own selection based upon the same criteria, and the names were forwarded to the researcher.

Once the sample had been selected, packets of thirty-six questionnaires with suitable cover letters (see Appendix B) were sent to the participating districts. Beginning approximately two weeks later, the researcher visited the districts to collect the instruments and leave follow-up instruments for the nonrespondents. Because of the end of the school year, and a fairly high return rate, no additional follow-up was made.

CHAPTER IV. ANALYSIS OF DATA AND FINDINGS

Analysis of Data

The data obtained from this study was analyzed using SPSS procedures, FREQUENCIES and CROSSTABS. Statistical options used included the Chi Square Test of significance which is part of the CROSSTABS package. FREQUENCIES was used primarily to tabulate raw data and per cent of response from the questionnaires, with selected groupings including teachers and administrators, grade level within teachers, and three groups of school districts established by the use of the McNally criteria mentioned in Chapter III. CROSSTABS was used to compare responses to questions by position (teachers and administrators) and to compare responses to items by grade level within the set of teachers responding. CROSSTABS was also used to compare the three groups of school districts on many of the items, and to compare within the groups of school districts by position and by grade level within teachers as described above. In tables where differences appeared to be large, the Chi Square Test was used. In several cases, some regrouping was required in order to have expected cell frequencies of at least five.

The multiple response questions (numbers five through nine) were tallied using FREQUENCIES. The results were summarized across each criteria category by counting the total number of responses regardless of position as well as by calculating a weighted score by counting three for each first response, two for each second response, and one for each third response. The scores obtained both ways were used to rank the criteria within each category.

The data from the responses recorded on computer scan sheets were also processed by using FREQUENCIES, with results tallied across the entire sample and subdivided by position, with per cent of responses noted.

Findings

Representativeness of the sample

The initial sample involved a total of 324 participants selected at random with the restriction that from each of the nine districts surveyed there would be selected fifteen elementary teachers, fifteen secondary teachers, three elementary building administrators, and three secondary building administrators. The study had a total return of 239

Table 4. Sample response: position by grade level

Position	Grade Level		
	Elementary	Secondary	Total
Principal	22	23	45
Teacher	102	92	194
Totals	115	124	239

instruments which were at least partly usable, for a return rate of 73.8 per cent. The breakdown of the sample shown in Table 4 is not different from the original scheme in any practical sense, since per cent of sample by grade level (whether principals, teachers, or total) never

varies more than 2.6 per cent from 50 per cent. The return by administrators was somewhat higher than that of teachers (83.3 per cent to 72.2 per cent), thus decreasing the ratio of teachers to administrators from 5 to 1 to about 4.3 to 1. Since the sample of administrators was small initially, this did not appear to have much impact, especially when teachers were also considered separately throughout this chapter.

When teaching experience was analyzed, it also appeared that we had a fairly representative situation. Table 5 shows experience within the districts tabulated by position, and by grade level within position. Even though more secondary teachers have less experience, this is not statistically significant (Chi Square = 8.3 with 4 degrees of freedom). Also when comparing total experience across position and grade level

Table 5a. Experience within districts by position and grade level within position (per cent)

Experience (%)	Principals			Teachers		
	Ele.	Sec.	Tot.	Ele.	Sec.	Tot.
1-2 years	13.0	4.5	8.9	7.8	16.5	11.5
3-5 years	4.3	18.2	11.1	17.6	19.8	18.7
6-10 years	34.8	27.3	31.1	29.4	14.3	22.3
11-15 years	4.3	18.2	11.1	19.6	22.0	20.7
over 15 years	43.5	31.8	37.8	23.5	27.5	26.4
Total number	23	22	45	102	92	104

within position (Table 5b), we note that a majority of administrators and a clear plurality of teachers have over fifteen years of experience.

Table 5b. Total experience by position and grade level within position (per cent)

Experience (%)	Principals			Teachers		
	Ele.	Sec.	Tot.	Ele.	Sec.	Tot.
1-2 years	0.0	0.0	0.0	4.9	9.9	7.3
3-5 years	0.0	0.0	0.0	13.7	14.3	14.0
6-10 years	13.6	30.4	22.2	24.5	16.5	20.7
11-15 years	18.2	4.3	11.1	14.7	14.3	14.5
over 15 years	68.2	65.2	66.7	42.2	45.1	43.5
Total number	22	23	45	102	92	194

The total results compare favorably with statistics from the State Department of Public Instruction which indicated the following percentages for its staff.

Table 6. Experience of teachers and principals statewide

Response	Teacher	Principal
1-2 years	10.6	0.0
3-5 years	17.6	3.3
6-10 years	24.3	11.4
11-15 years	15.3	19.6
over 15 years	32.1	65.6
Total number	31,202	1,350

The tables indicate that our sample is slightly older among teachers and somewhat younger among principals than state average. Neither is un-

expected since larger school systems tend to employ more specialized teachers and also compensate better for upper levels of experience, while offering more entry level administration positions in order to staff a larger number of schools. There does not appear to be any practical difference between our sample and the state figures which should be harmful to this study.

Evaluative criteria

Respondents to the survey were asked to consider a list of 155 statements which could be and have been used as evaluative criteria in a variety of school districts. The criteria were divided into five general categories: Personal Characteristics, Professional Qualities, Classroom Activities, Interpersonal Relations, and Product Measures. The teachers and administrators were asked to consider each statement and respond to the question, "Is this a standard by which teachers are evaluated in my school?" The purpose of this section of the instrument was to determine what criteria were perceived to be in use in the schools surveyed. The responses to the question were:

A. You know that the statement is definitely used as a standard.

B. You feel that the statement is probably used as a standard.

C. You feel that the statement is probably not used as a standard.

D. You know that the statement is definitely not used as a standard.

E. You have no opinion.

Appendix C lists the criteria along with the following information: per cent of the total number of respondents selecting each response and the number of respondents, per cent of teachers giving each response and

the number of teachers responding, and the per cent of administrators responding to each answer with the total number of administrators responding. The following criteria, listed by category, received at least 90 per cent of the total vote that they either definitely or probably are used as standards:

Personal characteristics

3. The teacher is physically and emotionally able to perform required duties (93.5 per cent of total, 94.1 per cent of teachers, 90.0 per cent of principals).
5. The teacher has a positive, enthusiastic attitude (90.1 per cent of teachers).
6. The teacher meets classroom situations with a sense of humor (90.0 per cent of principals).
7. The teacher demonstrates consistency and reliability as well as flexibility and adaptability (91.8 per cent of total, 91.8 per cent of teachers, 92.5 per cent of principals).
9. The teacher is self-confident (95.0 per cent of principals).
11. The teacher is reliable (93.1 per cent of total, 93.2 per cent of teachers, 92.5 per cent of principals).
14. The teacher possesses a business-like or task-oriented behavior (90.0 per cent of principals).

Professional Qualities

25. The teacher demonstrates a positive attitude toward teaching (92.2 per cent of total, 93.2 per cent of teachers).
29. The teacher accepts responsibilities such as hall duty, extra-curricular duties, service on committees, etc. (92.5 per cent of principals).
33. The teacher complies with the rules and regulations of the school system (93.5 per cent of total, 94.3 per cent of teachers).

Classroom Activities

50. The teacher has written objectives (95.0 per cent of principals).
51. The teacher prepares sufficient and appropriate written lesson plans for himself/herself or a substitute (95.0 per cent of principals).
52. The teacher develops plans consistent with the short and long range goals and objectives of the course (90.0 per cent of principals).

53. The teacher uses appropriate activities and resources to meet the needs of varied student or group abilities and interests (90.1 per cent of total, 90.1 per cent of teachers, 92.5 per cent of principals).
57. The teacher provides for individual differences (91.3 per cent of total, 90.6 per cent of teachers, 95.0 per cent of principals).
58. The teacher is aware of individual differences (93.9 per cent of total, 94.8 per cent of teachers, 90.0 per cent of principals).
62. The teacher exhibits a mastery of subject matter appropriate for the grade or skill level taught (94.4 per cent of total, 94.2 per cent of teachers, 95.0 per cent of principals).
63. The teacher knows and uses content which is appropriate to the subject area (91.3 per cent of total, 91.6 per cent of teachers).
65. The teacher uses content appropriate to the students' abilities and needs (92.6 per cent of total, 92.1 per cent of teachers, 95.0 per cent of principals).
67. The teacher develops lessons and makes assignments which are clear and consistent with course goals (97.5 per cent of principals).
68. The teacher makes clear, concise explanations (92.5 per cent of principals).
69. The teacher gives clear directions (92.3 per cent of principals).
70. The teacher relates current lessons to previous learning (90.0 per cent of principals).
75. The teacher involves students (94.4 per cent of total, 95.3 per cent of teachers, 90.0 per cent of principals).
78. The teacher uses praise and positive reinforcement (92.3 per cent of principals).
84. The teacher is well-organized but flexible (91.8 per cent of total, 91.1 per cent of teachers, 95.0 per cent of principals).
86. The teacher exhibits a friendly but positive control of the class (95.2 per cent of total, 95.8 per cent of teachers, 92.5 per cent of principals).
87. The teacher is firm, consistent, and fair (94.3 per cent of total, 95.3 per cent of teachers, 90.0 per cent of principals).
90. The teacher supports building and district discipline policies (90.5 per cent of total, 90.1 per cent of teachers, 92.5 per cent of principals).
103. The teacher makes the classwork interesting - puts his/her material across in an interesting way (92.1 per cent of principals).
104. The teacher conducts a classroom in which pupils actively participate in classroom discussion and activities (94.7 per cent of principals).

Interpersonal Relations

114. The teacher is available to students and offers additional assistance (92.1 per cent of principals).
115. The teacher is fair, impartial, and objective in his/her treatment of students (92.1 per cent of principals).
119. The teacher is understanding, encouraging, and helpful (93.4 per cent of total, 93.1 per cent of teachers, 94.7 per cent of principals).
122. The teacher has the respect of the students (91.6 per cent of total, 93.7 per cent of teachers).
123. The teacher has a good rapport with students (91.2 per cent of total, 90.4 per cent of teachers, 94.7 per cent of principals).
134. The teacher complies with authorized policies and procedures (92.5 per cent of total, 93.1 per cent of teachers).
135. The teacher is willing to accept his share of building responsibilities (90.7 per cent of total, 90.5 per cent of teachers, 92.1 per cent of principals).
137. The teacher deals effectively with parents (90.3 per cent of total, 92.1 per cent of principals).

Product Measures

None of these received 90 per cent of the A and B responses.

In addition to these criteria which were highly rated, it is interesting to note those which did not receive a majority of A or B responses. For the category of Personal Characteristics, these were criteria 8, 12, 15, and 16 overall and among teachers, but only criterion 15 among administrators. For Professional Qualities, criteria 17, 19, 21 and 31 did not receive a majority overall, while 19, 21, 31, and 36 failed to receive a majority among teachers, and 17, 19, 21, and 31 did not obtain a majority of A and B responses from principals. In the category of Classroom Activities, criteria 83, 92, 100, and 108 did not receive a majority from the total sample or from the teachers. None of the criteria in this category failed to receive a majority from the principals. For Interpersonal Relations, none of the criteria failed to receive a majority vote on responses A and B. Finally, in Product Measures, only 141 and

154 did receive a majority of the responses A and B in the total sample and among teachers, with 144 and 145 joining 141 and 154 for administrators.

When participants were asked to rank the three most important criteria within each category, additional information was gathered (see Appendix D for the complete listing and two sets of ranking, one based upon frequency of the response and the other a score which weighted the responses by position). The top five criteria within each category are basically unchanged by the method used to rank them, and are listed below, in order by category.

Personal Characteristics

Rank	Number	
1.	5.	The teacher has a positive, enthusiastic attitude.
2.	7.	The teacher demonstrates consistency and reliability as well as flexibility and adaptability.
3.	3.	The teacher is physically and emotionally able to perform required duties.
4.	9.	The teacher is self-confident.
5.	11.	The teacher is reliable.

Professional Qualities

1.	25.	The teacher demonstrates a positive attitude toward teaching.
2.	27.	The teacher assesses each lesson and unit in terms of student responses to techniques, activities, and materials, and in terms of student achievement of objectives.
3.	42.	The teacher is critical of and constantly trying to improve his/her work.
4.	26.	The teacher promotes the positive value of education.
5.	33.	The teacher complies with the rules and regulations of the school system.

Classroom Activities

1.	53.	The teacher uses appropriate activities and resources to meet the needs of varied student or group abilities and interests.

Rank	Number	
2.	51.	The teacher prepares sufficient and appropriate written lesson plans for himself/herself or a substitute.
3.	57.	The teacher provides for individual differences.
4.	62.	The teacher exhibits a mastery of subject matter appropriate for the grade or skill level taught.
5.	52.	The teacher develops plans consistent with the short and long range goals and objectives of the course.

Interpersonal Relations

1.	111.	The teacher shows respect for and interest in all pupils as individuals.
2.	115.	The teacher is fair, impartial, and objective in his/her treatment of students.
3.	119.	The teacher is understanding, encouraging, and helpful.
4.	123.	The teacher has a good rapport with the students.
5.	121.	The teacher makes each child feel important and respected.

Product Measures

1.	141.	The teacher is responsible for class achievement of course objectives.
2.	154.	The teacher is responsible for developing good citizenship in pupils.
3.	144.	The teacher is responsible for students' attitudes toward his/her class.
4.	153.	The teacher is responsible for developing social responsibility in pupils.
5.	145.	The teacher is responsible for students' attitudes toward his/her subject.

Analyses of Questions 10 through 40

Table 7. Question 10 responses

Of the five categories of statements listed below, which area do you feel is most frequently used as the primary source of evaluative criteria in your school? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Personal Characteristics	7.1	7.1	7.1	7.6	5.0	4.4	1.4	14.5
Professional Qualities	16.4	17.2	20.9	18.9	5.0	14.7	20.3	14.5
Classroom Activities	60.9	56.6	58.1	57.3	77.5	57.4	64.9	60.2
Interpersonal Relations	12.4	15.2	10.5	13.0	10.0	14.7	12.2	10.8
Product Measures	3.1	4.0	2.3	3.2	2.5	8.8	1.4	0.0
Total Responses (freq.)	225	99	86	185	40	68	74	83

From the data in Table 7 it is clear that the most popular category of evaluative criteria is classroom activities, being selected nearly four times as often by the respondents as any other category. It is particularly interesting to note that it was selected by over 77 per cent of the administrators. There was no significant difference between the choices made by teachers and those made by administrators, nor between teachers divided by grade level as determined by a Chi Square test. When considering the data for the grouped districts, it is notable that

those in the lower third selected personal characteristics more than twice as often as the total sample, and the upper third selected product measures more than twice as often as the entire sample, and far more often than the lower group which did not select product measures at all.

Table 8. Primary category of evaluative criteria by district level

Primary Source	District Levels (Thirds)			
	Lower	Middle	Upper	Total
Product Measures and Personal Characteristics	12	2	9	23
Interpersonal Relations	9	9	10	28
Classroom Activities	50	48	39	137
Professional Qualities	12	15	10	37
Totals	83	74	68	225

Chi Square = 7.9 with 6 degrees of freedom, not significant for $\alpha < .05$

It can be noted that the Lower Third had twelve responses that indicated Personal Characteristics were the most frequently used criteria. This compares to one for the Middle Third and three for the Upper Third. Similarly, the Upper Third selected Product Measures six times, compared to zero for the Lower Third and one for the Middle Third. However, when Product Measures and Personal Characteristics were combined, and tabulated by district level (Table 8), Chi Square value of 7.9 was found which was not significant for $\alpha < .05$ and 6 degrees of freedom.

Table 9. Question 11 responses

Of the five categories listed in number 10, which area do you feel is considered least in making evaluative decisions? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Personal Characteristics	14.5	14.9	11.4	13.2	21.1	17.6	15.7	11.0
Professional Qualities	9.1	7.4	10.2	8.8	10.5	8.8	8.6	9.8
Classroom Activities	5.0	5.3	6.8	6.0	0.0	7.4	4.3	3.7
Interpersonal Relations	10.5	11.7	10.2	11.0	7.9	8.8	8.6	13.4
Product Measures	60.9	60.6	61.4	61.0	60.5	57.4	62.9	62.2
Total Responses (freq.)	220	94	88	182	38	68	70	82

The data from Question 11 clearly indicate that Product Measures are the least favored as a source of evaluative criteria, having been selected by 60.9 per cent of the total number of respondents. It is interesting to note that among teachers there is no unanimity on this point, but that all of the 38 administrators did not select Classroom Activities, which agrees with their choice of that category on Question 10. When comparisons were made between teachers and administrators, teachers by grade level, and districts by group, no significant differences were found using Chi Square.

Table 10. Question 12 responses

Have the criteria which are used to evaluate teachers in your school been carefully explained to the teachers? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Yes	73.5	70.0	67.3	68.4	97.5	85.7	70.7	65.9
No	26.5	30.0	33.3	31.6	2.5	14.3	29.3	34.1
Total Responses (freq.)	230	100	90	190	40	70	75	85

The responses to this question show that the majority of all groups perceive that evaluative criteria have been clearly explained to teachers. However, administrators are nearly unanimous on this point (97.5 per cent responding "yes"), while most other groups have affirmative response rates of about 70 per cent. When administrators are compared to teachers, this difference becomes increasingly apparent (Table 11). A Chi Square Test with one degree of freedom yielded a value of 12.9 which is significant at the 0.001 level.

Table 11. Response to question 12 by position

Criteria Explained	Position		
	Principal	Teacher	Total
Yes	39	130	169
No	1	60	61
Totals	40	190	230

Table 11 (continued)

Corrected Chi Square = 12.9 with 1 degree of freedom and significance for $\alpha < .001$

When teachers were compared by grade level, no significant difference was found. Considering districts by level, a significant relationship was found between district level and perceptions of the respondents (Table 12).

Table 12. Response to question 12 by district level

Criteria Explained	Districts by Level			
	Upper Third	Middle Third	Lower Third	Total
Yes	60	53	56	169
No	10	22	29	61
Total Responses	70	75	85	230

Chi Square = 8.2 with 2 degrees of freedom and significance for $\alpha < .05$

This can be seen in even greater contrast when the Upper Third is compared to the lower third. The Upper Third showed an 85.7 per cent affirmative response while the Lower Third had only 65.9 per cent. Also, when the district groups were further broken down by position, it was noted that the Lower and Middle Thirds showed much less agreement between teachers and administrators than did the Upper Third, with principals giving far more affirmative responses proportionally than did teachers. These results were not checked for significance due to low expected frequencies within one cell. A Chi Square test comparing district levels using teachers only had a Chi Square of 9.3 with 2 degrees of freedom which is significant for $\alpha < .01$ (Table 13).

Table 13. Response to question 12 by district level - teachers only responding

Response (teachers only)	Districts by Level			
	Upper Third	Middle Third	Lower Third	Total
Yes	50	39	41	130
No	10	22	28	60
Total Responses	60	61	69	190

Chi Square = 9.3 with 2 degrees of freedom and significance for $\alpha < .01$

Table 14. Question 13 responses

Were teachers involved in the selection of evaluative criteria? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Yes	61.3	56.5	59.5	58.0	75.6	80.0	72.2	63.7
No	38.7	43.5	40.5	42.0	24.4	20.0	27.8	36.3
Total Responses (freq.)	217	92	84	176	41	65	72	80

When asked about the involvement of teachers in the selection of evaluative criteria, some interesting features were noted. Table 14 reveals that a higher proportion of administrators who responded perceived teacher involvement than did the participating teachers. This difference, however, was not statistically significant. A greater contrast was noticed when comparing district levels.

Table 15. Response to question 13 by district level

Teachers involved in selecting criteria?	Districts by Level			
	Upper Third	Middle Third	Lower Third	Total
Yes	52	52	29	84
No	13	20	51	133
Total Responses	65	72	80	217

Chi Square = 34.4 with 2 degrees of freedom and significance for $\alpha <$ .0001.

Table 15 shows a highly significant relationship between district level and teacher involvement in the selection of evaluative criteria, with districts in the Upper Third reporting teacher involvement nearly twice as often as those in the Lower Third. It is also noteworthy that when the Middle Third is broken down by position, all the administrators claimed that teachers had been involved in the selection of evaluative criteria while only 64.9 per cent of the teachers responded in the affirmative.

Table 16. Question 14 responses

Are the criteria written? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Yes	90.8	91.9	86.5	89.4	97.6	90.0	89.2	92.9
No	9.2	8.1	13.5	10.6	2.4	10.0	10.8	7.1
Total Responses (freq.)	229	99	89	188	41	70	74	85

There is almost unanimous agreement across all levels that school districts have written evaluative criteria. It is surprising that there is not a unanimous response among administrators, but the 2.4 per cent represents only one negative response.

Table 17. Question 15 responses

Are administrators supplied with guidelines for the uniform application of the criteria during evaluation? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Yes	87.4	85.9	87.8	86.8	90.0	93.8	92.9	76.7
No	12.6	14.1	12.2	13.2	10.0	6.3	7.1	23.3
Total Responses (freq.)	207	85	82	167	40	64	70	73

There appears to be high agreement across position and grade level within teachers that administrators have guidelines for the uniform application of evaluative criteria. However, there is some question whether there is agreement across the levels of districts. Table 18 shows the responses to Question 15 tabulated by district level. There is a highly significant relationship between District Level and perceived existence of administrative guidelines, with the Lower Third indicating the most negative response. When the Lower Third is examined by considering only teachers, there are still 14 negative responses, and only 43 positive responses. When teacher responses are tabulated by district

Table 18. Response to question 15 by district level

Do administrators have guidelines?	Districts by Levels			
	Upper Third	Middle Third	Lower Third	Total
Yes	60	65	56	181
No	4	5	17	26
Total Responses	64	70	73	207
Chi Square = 11.8 with 2 degrees of freedom and significance for $\alpha < .01$				

level, we note another significant relationship (Table 19). Similar results using administrators cannot be tested using Chi Square due to expected cell frequencies less than 5.

Table 19. Responses to question 15: teachers by district level

Responses	Teachers by District Level			
	Upper Third	Middle Third	Lower Third	Total
Yes	50	52	43	145
No	4	4	14	22
Total Responses	54	56	57	167
Chi Square = 9.8 with 2 degrees of freedom and significance for $\alpha < .01$				

Table 20. Question 16 responses

Are the same criteria (standards) used for all teachers? (adjusted frequencies, per cent response)

Responses (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Yes	87.6	83.5	91.5	87.3	88.9	82.3	99.4	91.0
No	12.4	16.5	8.5	12.7	11.1	17.7	11.6	9.0
Total Responses (freq.)	209	91	82	173	36	62	69	78

From the data presented in Table 20, it is clear that the same criteria are seen to apply to all teachers in nearly 90 per cent of the responses. This is true regardless of position, grade level of teachers, or level of district on the McNally criteria. No significant differences were found. It is of some interest that the Upper Third districts had the lowest rate of affirmative response, but when compared to the other two levels, there was no significant difference.

Table 21. Question 17 responses

Should the same criteria be used to evaluate all teachers? (adjusted frequencies, per cent response).

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Yes	74.4	79.6	64.6	72.6	83.3	71.9	74.6	76.3

Table 21 (continued)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
No	25.6	20.4	35.4	27.4	16.7	28.1	25.4	23.7
Total Responses (freq.)	211	93	82	175	36	64	71	76

It would appear that there is substantial agreement that the same standard ought to be applied to all teachers. The major source of disagreement is between teachers of different grade levels (Table 22). Here we find that secondary teachers are less likely to agree that the same standards should be applied to all teachers. The same results were found when the teachers of the Upper Third were compared by grade level. The remaining levels were not shown to be significant.

Table 22. Responses to question 17: teachers by grade level

Response	Teachers		
	Elementary	Secondary	Total
Yes	74	53	48
No	19	29	127
Total Responses	82	93	175

Corrected Chi Square = 4.2 with 1 degree of freedom and significance for $\alpha < .05$

Table 23. Question 18 responses

Are the criteria used in your evaluations based upon "Job Targets" or some kind of management by objectives approach? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Yes	64.5	69.6	57.1	63.8	67.6	89.7	60.3	46.3
No	35.5	30.4	42.9	36.2	32.4	10.3	39.7	53.7
Total Responses (freq.)	183	79	70	149	34	58	58	67

The data reported in Table 23 indicate that some sort of management by objectives is used by nearly two-thirds of all the persons surveyed.

There is no significant difference in the reported use of "Job Targets" when comparing by position or by grade level within teachers. However, when districts are compared across levels of agreement with the McNally criteria (Table 24), we have a highly significant relationship.

Table 24. Responses to question 18 by district level

Are "Job Targets" used?	District Levels			
	Upper Third	Middle Third	Lower Third	Total
Yes	52	35	31	118
No	6	23	36	65
Total Responses	58	58	67	183

Chi Square = 26.2 with 2 degrees of freedom and significance for $\alpha < .0001$

The table appears to indicate that the districts in the Upper Third are far more likely to use management by objectives than those in the Lower Third. It can also be noted that the lower frequency of response to this question, especially in the Middle and Lower Thirds may be due to an unfamiliarity with the term "Job Targets."

Table 25. Question 19 responses

Should teachers be involved in determining what criteria will be used in the evaluation of their teaching? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Yes	95.8	99.0	94.0	96.7	91.7	96.9	97.2	93.8
No	4.2	1.0	6.0	3.3	8.3	3.1	2.8	6.2
Total Response (freq.)	216	96	84	180	36	64	71	81

There was a very strong agreement across all levels listed that teachers should be involved in the selection of their evaluative criteria. This was especially true among elementary teachers as indicated by their 99 per cent affirmative response. The results here can be contrasted with those for Question 13 displayed in Table 14.

Table 26. Question 20 responses

Are teacher evaluations in your building relatively free from personal biases of the evaluator? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Yes	69.8	70.6	59.0	65.0	91.7	67.8	71.9	69.7
No	30.2	29.4	41.0	35.0	8.3	32.2	28.1	30.3
Total Responses (freq.)	199	85	78	163	36	59	64	76

The data indicate that about 30 per cent of those surveyed felt that there was personal bias present in the evaluation process. This reflected a 35 per cent response of teachers, but a mere 8.3 per cent response from administrators. Table 27 shows this relationship which resulted in a

Table 27. Responses to question 20 by position

Are evaluations relatively free from evaluator bias?	Position		
	Teachers	Principals	Total
Yes	106	33	139
No	57	3	60
Total Responses	163	36	199

Corrected Chi Square = 8.7 with 1 degree of freedom and significance for $\alpha < .01$

highly significant Chi Square of 8.7 for 1 degree of freedom. Secondary teachers gave a higher proportion of negative responses than elementary

teachers, but this was not a significant difference. The greatest disparity between teachers' and administrators' perceptions was seen in the Lower Third of districts where all 15 responding administrators felt that evaluations were relatively bias free, and only 62.3 per cent of the teachers agreed. There was no significant difference on the presence of bias across district levels, however.

Table 28. Question 21 responses

Is a person who is involved with a popular extra-curricular activity more likely to be evaluated on the success of that activity than on his/her teaching? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Yes	42.7	48.6	52.0	50.3	11.1	40.7	41.9	44.9
No	57.3	51.4	48.0	49.7	88.9	59.3	58.1	55.1
Total Responses (freq.)	185	74	75	149	36	54	62	69

The results of Question 21 were quite clear. First, from the overall response, it would seem that a substantial minority of subjects believe that success in a popular extra-curricular activity does affect teacher evaluation. Much more striking was the comparison of teachers with administrators (Table 29). While teachers were almost evenly divided on the question, nearly 90 per cent of the building administrators reported that extra-curricular activities were not likely to supplant

Table 29. Response to question 21 by position

Response	Position		
	Teacher	Principal	Total
Yes	75	4	79
No	74	32	106
Total Responses	149	36	185

Corrected Chi Square = 16.7 with 1 degree of freedom, significant for $\alpha < .0001$.

normal teacher evaluation based upon teaching. The Chi Square Test revealed a highly significant result of 16.7 with 1 degree of freedom which gives a level of significance less than .0001. A similar result was indicated within district level for the Upper and Middle Thirds ($\alpha < .05$), with the Lower Third narrowly missing significance at the .05 level.

Table 30. Question 22 responses

From the list of possible teacher evaluators, select the letter of the person(s) who serve as the primary evaluator in your building.

Response (%)	All	Position				Districts by Level		
		Teacher			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Principal or Assistant Principal	93.1	98.9	84.7	92.2	97.2	87.5	91.7	98.7
Supervisor or Curriculum Specialist	0.9	1.1	1.2	1.1	0.0	0.0	2.8	0.0
Department Head	2.8	7.1	0.0	3.3	0.0	9.4	0.0	0.0

Table 30 (continued)

Response (%)	All	Position				Districts by Level		
		Teacher			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Peers (fellow teachers)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Students	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Central Office Personnel	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Others (Specify)	3.2	0.0	7.1	3.3	2.8	3.1	5.6	1.2
Total Responses (freq.)	216	95	85	180	36	64	72	80

The responses to Question 22 clearly indicated that the primary evaluator continues to be the building principal and his/her assistant. When the "others" category was tabulated, it was noted that every response included the principal and either a supervisor or department head. Noting this, it is apparent that the administrators were unanimous, and all other categories exceeded 95 per cent of the responses. The only contrast came when teachers were compared by grade level. Here it appears that secondary teachers have multiple evaluators, something not common at the elementary level since the building administrator is the only person with the authority and responsibility to evaluate.

Table 31. Question 23 responses

Of those personnel listed in number 22, who do you think is best equipped to evaluate teachers? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Principal or Assistant Principal	57.3	62.8	45.8	54.8	69.4	54.7	52.1	64.1
Supervisor or Curriculum Specialist	6.1	3.2	8.4	5.6	8.3	0.0	11.3	6.4
Department Head	9.4	3.2	18.1	10.2	5.6	18.8	5.6	5.1
Peers (fellow teachers)	14.6	21.3	12.0	16.9	2.8	9.4	18.3	15.4
Students	1.2	0.0	3.6	1.7	2.8	1.6	4.2	0.0
Central Office Personnel	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Others (Specify)	10.8	9.6	12.0	10.7	11.1	15.6	8.5	9.0
Total Responses (freq.)	213	94	83	177	36	64	71	78

When participants were asked to select the person or persons they thought were the best evaluator(s), the results were not nearly as overwhelming as on Question 22. The principal or assistant principal was favored by the majority in all categories except secondary teachers. The administrators themselves favored principal evaluations by nearly 70 per cent, and when the "other" responses from administrators which listed the principal and some other evaluator are included, this per cent jumps to more than 80. In most of the "other" cases, the principal was listed

as one of several evaluators. Other combinations listed students and department heads or supervisors. One respondent called for the use of an outside evaluator. When comparing teachers by grade level, we have a clear divergence of opinion, but due to the small expected frequencies in several of the cells, and the general loss of any practical information when they are combined in a way to make a Chi Square Test reasonable, no test of significance was made. One item did stand out, however; the total agreement of all respondents that central office personnel not be involved in the evaluation of teachers.

Table 32. Question 24t responses: teachers only responding

How many times have you been formally observed this past year? (adjusted frequencies, per cent response)

Response (%)	Teachers			Districts by Levels		
	ele.	sec.	tot.	Upper Third	Middle Third	Lower Third
0	32.6	29.4	31.1	38.2	47.5	10.6
1	16.8	27.1	21.7	16.4	22.2	25.8
2	25.3	17.6	21.7	14.5	15.3	33.3
3-4	22.1	21.2	21.7	27.3	8.5	28.8
5 or more	3.2	4.7	3.9	3.6	6.8	1.5
Total Responses (freq.)	95	85	180	55	59	66

The information from responding teachers on the frequency of formal observations was somewhat more difficult to discuss. On the surface, it can be noted that over 50 per cent of all teachers responding reported at most one formal evaluation in the past year, with over 30 per cent re-

porting no formal evaluation at all. When the data were displayed by districts, it appears that there is some relationship between district levels and frequency of formal evaluation. Table 33 displays this information in raw form, with the last two categories combined to obtain expected frequencies of at least 5. The resulting Chi Square test value of 25.6 is highly significant ($\alpha < .001$), confirming a relationship between the district levels and the frequency of evaluation.

Table 33. Responses to question 24t by district level

Responses	District Level			
	Upper Third	Middle Third	Lower Third	Total
0	21	28	7	56
1	9	13	17	39
2	8	9	22	39
3 or more	17	9	20	46
Total Responses	55	59	66	180

Chi Square = 25.6 with 6 degrees of freedom and significance for $\alpha < .001$

Those who had responded that they had not been observed in the past year were also asked to indicate when they had last been formally observed. Most of the teachers indicated one or two years ago, and often this was due to the cyclical nature of their evaluation system. However, in one district which was part of the Middle Third, teacher estimates of when the last formal evaluation had taken place ranged from last year to twelve years ago! One teacher from that system claimed he/she had had no formal evaluation in the more than six years that he/she had been in the system. Another teacher indicated that he/she had never been evaluated

formally since joining a new system over three years ago and that he/she had last been evaluated six years ago in his/her former school district.

Table 34. Question 24a responses: administrators only responding

How frequently do you formally observe the typical teacher in your building? (adjusted frequencies, per cent response)

Response (%)	Administrators	Districts by Level		
		Upper Third	Middle Third	Lower Third
Three or more times a year	28.6	44.4	45.5	6.7
Twice a year	51.4	22.2	45.5	73.3
Annually	8.6	11.1	0.0	13.3
Semi-annually	0.0	0.0	0.0	0.0
Every three years or less	11.4	22.2	9.1	6.7
Total Responses (freq.)	35	9	11	15

The data for Question 24a are from such a very small sample that generalizations are more speculation. However, one point is quite clear; 80 per cent of the principals responding believe that they are involved in formal observation of the typical teacher twice or more times a year, and nearly 90 per cent believe that the typical teacher is observed at least annually. This is in contrast to the perceptions of the teachers as relfected by Table 32. Also, the table would seem to indicate <u>increased</u> frequency of multiple evaluations from the Lower Third to the Upper Third, which also is contrary to Table 32, and yet an increase of evaluation following a cycle of three years or more.

Table 25. Question 25t responses: teachers only responding

During your last formal observation, how long did the observer spend in your classroom? (adjusted frequencies, per cent response)

Response (%)	Teachers			Districts by Level		
	ele.	sec.	tot.	Upper Third	Middle Third	Lower Third
Less than 15 minutes	20.4	19.0	19.8	27.3	25.0	9.1
Between 15 and 30 minutes	34.4	28.6	31.6	29.1	30.4	34.8
Between 30 and 45 minutes	29.0	33.3	31.1	23.6	26.8	40.9
Between 45 and 60 minutes	10.8	15.5	13.0	16.4	8.9	13.6
Over 60 minutes	5.4	3.6	4.5	3.6	8.9	1.5
Total Responses (freq.)	93	84	177	55	56	66

Of the teachers responding, slightly more than half reported that their last formal evaluation observation lasted thirty minutes or less. There is no significant difference when comparing grade levels or district levels. It is interesting to note that teachers in the Lower Third reported proportionately fewer short observations than either of the other levels, while the Upper Third reported the most short observations proportionately.

Table 36. Question 25a responses: administrators only responding

How long is your average observation? (adjusted frequencies, per cent response)

Response (%)	Administrators	Districts by Level		
		Upper Third	Middle Third	Lower Third
Less than 15 minutes	8.6	0.0	27.3	0.0
Between 15 and 30 minutes	22.9	33.3	27.3	13.3
Between 30 and 45 minutes	45.7	11.1	45.4	66.7
Between 45 and 60 minutes	22.9	55.6	0.0	20.0
Over 60 minutes	8.6	0.0	0.0	0.0
Total Responses (freq.)	35	9	11	15

As in Table 34, the sample size is so small that inferences are not justified. Some observations seem appropriate, however. As a group, the principals who responded generally indicated that the typical formal observation lasted for 30 minutes or more as represented by a 77.2 per cent response. Also, all administrators in the Upper Third as well as the Lower Third admit to an average formal observation of at least 15 minutes. Both of these perceptions appear to differ with those of the teachers as shown in Table 35.

Table 37. Question 26t responses: teachers only responding

Do you feel that this (sample of your behavior) was adequate to make a judgment regarding your ability as a teacher? (adjusted frequencies, per cent response)

Response (%)	Teachers			Districts by Level		
	ele.	sec.	tot.	Upper Third	Middle Third	Lower Third
Yes	58.2	50.0	54.3	54.5	50.0	57.8
No	41.8	50.0	45.7	45.5	50.0	42.2
Total Responses (freq.)	91	82	173	55	54	64

The data displayed in Table 37 reveal that a majority of teachers responding feel that the amount of their teaching behavior which is observed formally is sufficient for making evaluative judgments. However, it is clear that a substantial minority (45.7 per cent) dissent on this point. There is no significant difference across district levels.

Table 38. Question 26a responses: administrators only responding

Do you feel that this (the amount of time spent on a typical observation) is adequate for making sound judgments? (adjusted frequencies, per cent response)

Response (%)	Administrators	Districts by Level		
		Upper Third	Middle Third	Lower Third
Yes	65.7	77.8	45.5	73.3
No	34.3	22.2	54.4	26.7
Total Responses (freq.)	35	9	11	15

The data again need little explanation except to note that a clear majority of administrators responding indicated that their typical evaluation observation provided an adequate sample of teacher behavior. Overall, about one-third of the respondents dissented from this opinion. It is of particular interest that within the administrators of the Middle Third the position is reversed, with the negative response in the majority. The remaining two levels voted for the affirmative position by a margin of about three to one. Sample size was too small to check for significance.

Table 39. Question 27 responses

Below are listed a variety of means for collecting evaluative data. Select the letter which proceeded the method most frequently used in your building. (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Administrative rating forms	73.8	--	--	62.5	76.5	80.0	73.3	70.6
Student rating forms	0.0	--	--	0.0	0.0	0.0	0.0	0.0
Systematic Observation	21.4	--	--	25.0	20.6	20.0	13.3	29.4
Self-evaluation	2.4	--	--	12.5	0.0	0.0	6.7	0.0
Video or audio tape	0.0	--	--	0.0	0.0	0.0	0.0	0.0
Student achievement	0.0	--	--	0.0	0.0	0.0	0.0	0.0
Peer ratings	0.0	--	--	0.0	0.0	0.0	0.0	0.0
Other (specify)	2.4	--	--	0.0	2.9	0.0	6.7	0.0

Table 39 (continued)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Total Responses (freq.)	42	--	--	8	34	10	15	17

This question failed to produce usable results except from administrators due to a flaw in the instrument which caused most teachers to skip past it as it appeared to be included in a section for administrators only. The data indicate that administrative ratings and systematic evaluation play the dominant role as the means of gathering evaluative data, at least as perceived by the participating administrators. Little else can be said.

Table 40. Question 28 responses

In teacher evaluation, is peer evaluation used? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Frequently	2.4	3.3	1.3	2.4	2.8	3.4	4.4	0.0
Sometimes	20.0	18.9	19.0	18.9	25.0	36.2	19.1	8.9
Never	77.6	77.8	79.7	78.7	72.2	60.3	76.5	91.1
Total Responses (freq.)	205	90	79	169	36	58	68	79

The data for Question 28 show that in the great majority of cases, peer evaluation is not used (77.6 per cent of total responses). There is no significant difference between teachers and administrators, nor between teachers by grade level. However, when the responses are considered by district level, combining the first two categories for the sake of expected cell frequencies, a highly significant relationship was revealed (Table 41, $\alpha < .001$ when Chi Square = 18.3 with 2 degrees of freedom). The data appear to indicate that this is due to an almost

Table 41. Responses to question 28 by district level

Responses	Districts by Levels			
	Upper Third	Middle Third	Lower Third	Total
Frequently or Sometimes	23	16	7	46
Never	35	52	72	159
Total Responses	58	68	79	205

Chi Square = 18.3 with 2 degrees of freedom and significance for $\alpha < .001$

total lack of peer evaluation in schools in the Lower Third, and increasing usage of peer evaluation from the Middle Third to the Upper Third. It should be noted that peer evaluation does not receive more than 40 per cent of the responses of "Frequently" or "Sometimes" even at the Upper Third.

Table 42. Question 29 responses

In teacher evaluation, is self-evaluation used? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Frequently	32.5	34.4	31.3	32.9	30.6	49.2	32.4	20.3
Sometimes	37.9	37.8	33.8	35.9	47.2	44.1	44.1	27.8
Never	29.6	27.8	35.0	31.2	22.2	6.8	23.5	51.9
Total Responses (freq.)	206	90	80	170	36	59	68	79

In response to the use of teacher evaluation, the participants indicated that this method was used at least sometimes in over 70 per cent of the cases. There is no difference in this perception when the total is divided by position or when teachers are split by grade level. However, when districts were considered by levels, a disparity did exist. The Upper Third reported that self-evaluation was used at least sometimes on over 95 per cent of the returns. This contrasted with 76.5 per cent for the Middle Third and 48.1 for the Lower Third. When tabulated by raw scores (Table 43), the results were found to be highly significant using a Chi Square test with 4 degrees of freedom. A similar result was found when only teachers' responses were considered by District Levels.

Table 43. Responses to question 29 by district levels

Responses	Districts by Levels			
	Upper Third	Middle Third	Lower Third	Total
Frequently	29	22	16	67
Sometimes	26	30	22	78
Never	4	16	41	61
Total Responses	59	68	79	206

Chi Square = 36.6 with 4 degrees of freedom and significance for $\alpha <$ .0001

Table 44. Question 30 responses

In teacher evaluation, are students' ratings used? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Frequently	1.0	1.1	1.3	1.2	0.0	1.7	0.0	1.3
Sometimes	16.7	13.3	18.2	15.6	22.2	24.1	17.9	10.3
Never	82.3	85.6	80.5	83.2	77.8	74.1	82.1	88.5
Total Responses (freq.)	203	90	77	167	36	58	67	78

The responses to Question 30 indicated that students' ratings are seldom used in the evaluation of teachers. There was general agreement across all levels of all variables reported. It is of interest to note that secondary teachers indicated only slightly more frequent use than did elementary teachers, with administrators indicating an even higher rate

than teachers. All differences were not significant.

Table 45. Question 31 responses

In teacher evaluation, is student achievement used? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Frequently	8.1	11.5	7.9	9.8	0.0	8.9	4.7	10.3
Sometimes	50.5	47.1	46.1	46.6	68.6	53.6	43.8	53.8
Never	41.4	41.4	46.1	43.6	31.4	37.5	51.6	35.9
Total Responses (freq.)	198	87	76	163	35	56	64	78

A 58.6 per cent majority of those responding to Question 31 have indicated that student achievement is used at least occasionally in teacher evaluation. The teachers reveal no significant difference when divided by grade level, and districts did not significantly differ across district levels. When teachers' responses were compared to those of principals, a possibly significant relationship was indicated. However, when the data were regrouped to insure sufficient cell expected values, the results were not significant.

Table 46. Question 32 responses

Virtually every evaluation program mentions "informal observation." How great a role should this play in the total evaluation of a teacher? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Very important	23.6	27.5	17.3	22.7	27.8	28.3	24.6	19.0
Important	64.9	65.9	64.2	65.1	63.9	61.7	63.8	68.4
Relatively unimportant	9.6	5.5	14.8	9.9	8.3	10.0	8.7	10.1
Insignificant	1.9	1.1	3.7	2.3	0.0	0.0	2.9	2.5
Total Responses (freq.)	208	91	81	172	36	60	69	79

The tabulated values indicated that nearly 90 per cent of all those responding consider "informal observation" to play an important role in teacher evaluation. This is true across position, grade level within teachers, and district levels.

Table 47. Question 33t responses: teachers only responding

If you are a teacher, in your opinion does your evaluator have sufficient training to be an effective evaluator? (adjusted frequencies, per cent response)

Response (%)	Teachers			Districts by Level		
	ele.	sec.	tot.	Upper Third	Middle Third	Lower Third
Yes	65.9	59.8	63.0	76.5	62.7	52.4

Table 47 (continued)

Response (%)	Teachers			Districts by Level		
	ele.	sec.	tot.	Upper Third	Middle Third	Lower Third
No	14.3	19.5	16.8	11.8	16.9	20.6
Don't know	19.8	20.7	20.2	11.8	20.3	27.0
Total Responses (freq.)	91	82	173	51	59	63

The results of Question 33t appeared to indicate that a majority of teachers surveyed felt that their evaluator has had sufficient training to be a good evaluator. There were differences across grade levels, with elementary teachers a bit more positive and secondary a bit more negative, but these differences were not significant. Similarly, there were differences across district levels, with the Upper Third more positive and the Lower Third more negative. Again, these findings were not significant.

Table 48. Question 33a responses: administrators responses only

If you are an administrator, how much formal training do you have which you would say specifically prepared you to evaluate teachers? (adjusted frequencies, per cent response)

Response (%)	Administrators	Districts by Level		
		Upper Third	Middle Third	Lower Third
Less than 5 hours	19.4	11.1	25.0	20.0
5-8 hours	27.8	22.2	25.0	33.3
9-16 hours	19.4	11.1	25.0	20.0
16-40 hours	19.4	33.3	0.0	26.7

Table 48 (continued)

Response (%)	Administrators	Districts by Level		
		Upper Third	Middle Third	Lower Third
more than 40 hours	13.9	22.2	25.0	0.0
Total Responses (freq.)	36	9	12	15

The above data indicated that the amount of formal preparation of evaluators varied greatly, with preparation ranging from very little (less than five hours) to a considerable amount (in excess of forty hours). What the respondents perceived as formal preparation may have also varied considerably, so that the results lack meaning.

Table 49. Question 34 responses

How many hours of inservice training are devoted each school year to help teachers improve in specific areas of weakness as shown by their evaluations? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
0-1	61.8	75.6	67.5	71.6	19.4	60.0	59.0	65.3
2-4	14.1	11.5	9.1	10.3	30.6	16.4	14.8	12.0
5-8	14.1	9.0	15.6	12.3	22.2	14.5	11.5	16.0
9-16	3.7	2.6	1.3	1.9	11.1	1.8	4.9	4.0
Over 16	6.3	1.3	6.5	3.9	16.7	7.3	9.8	2.7
Total Responses (freq.)	191	78	77	155	36	55	61	75

More than three-quarters of those responding to Question 34 indicated that their districts spent four hours or less per year in inservice programs designed to aid teachers in areas of weakness as shown by their evaluations. When comparing teachers to administrators, though, a conflict was noted. Table 50 was reduced since low expected frequencies

Table 50. Responses to question 34 by position

Responses	Position		
	Teachers	Administrators	Total
0-1 hours	111	7	118
2-4 hours	16	11	27
5 or more hours	28	18	46
Total Responses	155	36	191

Chi Square = 32.4 with 2 degrees of freedom and significance for $\alpha < .001$

forced a combining of the last three categories in the responses. The results still indicated that there was a highly significant relationship between response to the question and position of the respondent (Chi Square = 32.4 with 2 degrees of freedom and significance for $\alpha < .001$). The table also would seem to indicate that teachers perceive far less inservice time devoted to improvement of weaknesses than do administrators. Tabular results similar to those in Table 50 were found for the Lower Third and Middle Third of the district levels, but significance testing was not done because of the distortions which would have been necessary to obtain proper expected cell frequencies. No significant difference was found within teachers by grade level or across district levels.

Table 51. Question 35 responses.

Have teachers been involved in determining what procedures will be used to evaluate their teaching? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Yes	63.7	58.0	60.3	59.1	83.3	82.5	73.8	38.2
No	36.3	42.0	39.7	40.9	16.7	17.5	26.2	61.2
Total Responses (freq.)	190	81	73	154	36	57	65	68

Of those responding to Question 35, 63.7 per cent indicated that teachers had been involved in determining the evaluation procedures while 36.3 per cent indicated that teachers had not been involved in determining the evaluation procedures. This, however, does not reveal much of the true picture. When teachers were compared to administrators, it was noted that 83.3 per cent of the administrators claimed that teachers had been involved, while less than 60 per cent of the teacher acknowledged teacher involvement. This can be better illustrated in tabular form (Table 52) where it has been noted that there was a significant relationship between position and perceptions of teacher involvement in determining evaluation procedures, with administrators far more positive in their responses than teachers. There was no significant difference across teachers by grade level, nor were there any significant differences within district levels across grade levels of teachers, nor across positions. The major difference was revealed when districts were

Table 52. Responses to question 35 by position

Responses	Position		
	Teachers	Administrators	Total
Yes	91	30	121
No	63	6	69
Total Responses	154	36	190

Corrected Chi Square = 6.4 with 1 degree of freedom and significance for $\alpha < .05$

considered across district levels. With the Upper Third reporting teacher involvement in 82.5 per cent of the responses, and the Middle Third in 73.8 per cent of the responses, it was found that the Lower Third reported involvement in only 38.2 per cent of the responses. When district levels were compared, a Chi Square value of 30.6 was produced which with 2 degrees of freedom has significance at the .0001 level. From the table below, it seems clear that the major difference is attributable to the Lower Third. A similar result was also found when only the teachers were compared across district levels, again producing significance for $\alpha < .0001$.

Table 53. Responses to question 35 by district levels

Responses	Districts by Levels			
	Upper Third	Middle Third	Lower Third	Total
Yes	47	48	26	121
No	10	17	42	69
Total Responses	57	65	68	190

Chi Square = 30.6 with 2 degrees of freedom and significance for $\alpha < .0001$

Table 54. Question 36 responses

Is the evaluation procedure itself evaluated and updated regularly? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Yes	69.8	61.0	71.1	66.0	86.1	67.3	75.8	66.7
No	30.2	39.0	28.9	34.0	13.9	32.7	24.2	33.3
Total Responses (freq.)	189	77	76	153	36	55	62	72

Approximately 70 per cent of those responding to Question 36 indicated that the evaluation procedures used in their districts were regularly evaluated and updated. This left a fair-sized minority who do not believe that this occurs. When teachers and administrators were compared (Table 55) it was found that while teachers gave a positive

Table 55. Responses to question 36 by position

Responses	Position		
	Teachers	Administrators	Total
Yes	101	31	132
No	52	5	57
Total Responses	153	36	189

Corrected Chi Square = 4.7 with 1 degree of freedom and significance for $\alpha < .05$

response of 66 per cent, prinicpals gave an affirmative response in excess of 86 per cent. This yielded a significant Chi Square value of

4.7 with 1 degree of freedom. All other comparisons across grade levels within teachers and across district levels were not significant. One interesting aside, in the Upper Third, all responding administrators indicated that their procedures were updated regularly, while less than 61 per cent of the teachers responded similarly.

Table 56. Question 37 responses

Are you satisfied with the evaluation procedures used in your building? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Yes	62.3	69.8	55.8	63.2	58.3	64.4	65.2	57.7
No	37.7	30.2	44.2	36.8	41.7	35.6	34.3	42.3
Total Responses (freq.)	199	86	77	163	36	59	69	71

When consideration was given to the general satisfaction of the respondents toward their evaluation procedures it was first noted that a sizeable minority (37.7 per cent) of those responding were not satisfied with the procedures used in their schools. What did appear to be surprising was the fact that administrators were generally more negative than were teachers, although not significantly so. In particular, within the Lower Third, 53.3 per cent of the 15 principals responding were negative in their response compared to 39.3 per cent of the 56 teachers. Generally there was agreement across district levels. The

only conflict which was significant was with teachers across grade levels. While this was not significant over all teacher responding, Table 57 indicates that it was significant within the Upper Third where 57.1 per cent of the secondary teachers who responded indicated dissatisfaction and only 24.1 per cent of the elementary teachers were not satisfied.

Table 57. Responses to question 37 by grade level of teachers in the Upper Third

Responses	Teachers		
	Elementary	Secondary	Total
Yes	22	9	31
No	7	12	19
Total Responses	29	21	50

Corrected Chi Square = 4.3 with 1 degree of freedom and significance for $\alpha < .05$

Table 58. Question 38 responses

Below is a list of some of the possible purposes which can be served by teacher evaluation. Write the letter which corresponds to the purpose which you see as the main purpose of teacher evaluation as practiced in your school. (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Renewal/ Nonrenewal	19.3	22.0	20.0	21.1	11.1	11.1	10.3	34.2
Staff Assignment	6.3	8.8	6.3	7.6	0.0	6.3	7.4	5.3

Table 58 (continued)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Staff Reduction	3.9	3.3	6.3	4.7	0.0	3.2	1.5	6.6
Improvement of instruction	62.3	58.2	55.0	56.7	88.9	69.8	69.1	50.0
Differentiated Pay	0.5	0.0	1.3	0.6	0.0	0.0	1.5	0.0
Other (specify)	7.7	7.7	11.3	9.4	0.0	9.5	10.3	3.9
Total Responses (freq.)	207	91	80	171	36	63	68	76

The most frequent response chosen as the current purpose for teacher evaluation was "Improvement of instruction." This was true across all levels of all categories. Also, of those choosing "Other," six teachers included "Improvement of instruction" with at least one of the other replies, most frequently "Renewal/Nonrenewal." Another five teachers gave the major purpose to be fulfilling contractual requirements. A more cynical response by one teacher indicated that teacher evaluation's main purpose was to create jobs. The response which ranked second in overall was "Renewal/Nonrenewal" which received 19.3 per cent of the responses (21.1 per cent among teachers and 11.1 per cent among principals). Administrators appeared to perceive evaluation in one of two ways: either as a means for improving instruction or else as a tool to be used in non-renewal. Improvement of instruction was far more popular (88.9 per cent

to 11.1 per cent, a ratio of 8 to 1). The only conflict appeared when districts were considered across district levels. Here it seemed that the Lower Third saw "Renewal/Nonrenewal" as a much more important purpose than the other levels, and "Improvement of instruction" proportionately of less importance. Because of the large number of categories, and since too much distortion would have resulted from combining categories, Chi Square was not used to determine significance if in fact there was a significant difference.

Table 59. Question 39 responses

Of those possible purposes listed below, select the letter of the purpose which you feel teacher evaluation ought to serve in your school. (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Renewal/ Nonrenewal	1.4	1.1	1.2	1.1	2.8	1.6	1.4	1.3
Staff Assignment	2.8	3.2	3.7	3.4	0.0	3.2	1.4	3.8
Staff Reduction	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Improvement of instruction	92.4	92.5	90.2	91.4	97.2	92.1	92.8	92.4
Differentiated Pay	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Other (specify)	3.3	3.2	4.9	4.0	0.0	3.2	4.3	2.5
Total Responses (freq.)	211	93	82	175	36	63	69	79

The results of Question 39 indicated overwhelmingly that teachers and administrators feel that the main purpose of teacher evaluation ought to be for improvement of instruction. In addition, those who responded "Other" unanimously included "Improvement of instruction" with their choice of purpose. Included with this choice was either nonrenewal (4 times), staff assignment (once) and differentiated pay (three times). There was no difference between groups by position, grade level within teachers, or across district levels.

Table 60. Question 40 responses

If you were asked to select a symbol to represent teacher evaluation, what would it be? (adjusted frequencies, per cent response)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
A. A yard stick	43.4	42.4	41.8	42.1	50.0	44.1	44.4	41.9
B. A guillotine	1.0	0.0	2.5	1.2	0.0	0.0	0.0	2.7
C. A stethoscope	25.5	23.5	21.5	22.6	40.6	28.8	20.6	27.0
D. A Peeping Tom	2.6	2.4	3.8	3.0	0.0	0.0	3.2	4.1
E. A final examination	7.1	10.6	5.1	7.9	3.1	2.8	9.5	6.8
F. A psychiatrist's couch	2.0	1.2	2.5	1.8	3.1	5.1	6.3	0.0
G. Numbers in a hat	7.7	7.1	10.1	8.5	3.1	5.1	11.1	6.8
H. A popularity pageant	10.7	12.9	12.7	12.8	0.0	16.9	4.8	10.8

Table 60 (continued)

Response (%)	All	Position				Districts by Level		
		Teachers			Admin.	Upper Third	Middle Third	Lower Third
		ele.	sec.	tot.				
Total Responses (freq.)	196	85	79	164	32	59	63	74

Question 40, responses A through F were suggested by Armstrong (1973). Responses G and H were suggested by Dr. Rex Thomas of Iowa State University who apparently felt that Armstrong had overlooked some possibilities. Of those responding, 18.4 per cent appeared to agree with Thomas. In general, it was noted that nearly 70 per cent of all respondents saw teacher evaluation symbolized by either a yard stick or a stethoscope, which are fairly neutral symbols. Another 7.1 per cent selected a final examination, a symbol which generally has a variety of meanings depending on the perceptions and experiences of the individual respondent. Another 2 per cent selected the psychiatrist's couch, a symbol which also has many possible meanings. This left 22 per cent of the responses for symbols which appear to have basically negative connotations. It is worthy of note that responses G and H drew 18.4 per cent of all observations, and 21.3 per cent of the responses from teachers. Teachers in general reflected more negativeness than did administrators with 25.5 per cent negative responses compared to only 3.1 per cent for administrators. On the other hand, principals viewed teacher evaluation in a much more neutral fashion by selecting A or C in 90.6 per cent of the responses. When the categories were combined by grouping responses of A and C into

one category and all other responses into another, and tabulating the results across positions (Table 61), it was found that a highly significant relationship existed between position and response category (Chi Square = 8.5 with 1 degree of freedom). There were no significant differences between grade levels within teachers, nor across district levels.

Table 61. Responses to question 40 by position

Responses	Position		
	Teachers	Administrators	Total
A, C	106	29	135
B, D, E, F, G, H	58	3	61
Total Responses	164	32	194
Chi Square = 8.5 with 1 degree of freedom and significance for $\alpha < .001$			

CHAPTER V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary of Findings

Preliminary information from the original twenty-five school districts which had been contacted by telephone indicated that formal evaluation in some form would seem to be part of the school policies in virtually every district in the state. Of the twenty-one districts which sent materials, nearly all had some kind of written evaluative criteria and all showed some form of written evaluation procedure. In most cases, the procedures involved the use of administrator ratings and little else. Overall, there was compliance with Iowa Statute regarding teacher evaluation, although there were some indications that teacher involvement in the determination of procedures was not universal.

The participating school districts were divided into three levels (Upper, Middle, and Lower Thirds) using criteria for a good evaluation program suggested by McNally. These levels were used to determine whether there was general agreement across districts regarding the purpose, criteria, and procedures used in teacher evaluation. Returns indicated that the primary purpose of evaluation was the improvement of instruction. There were no significant differences across district levels on this point. A secondary purpose appeared to concern renewal/nonrenewal of teacher contracts.

The districts were also in general agreement that the major category of evaluative criteria should be "Classroom Activities," with the use of "Product Measures" seen as least desirable. Several significant differences were found across district levels when the mechanics of applying evaluative

criteria were discussed. As the level of compliance with the McNally criteria increased, there was stronger perception among respondents that evaluative criteria had been carefully explained to teachers. This was supported not only by the perceptions of the teachers, but also when all participants were considered. There was a highly significant relationship across district levels concerning teacher involvement in selection of the evaluative criteria. A similar relationship involved the use of job targets, and the existence of administrative guidelines for the uniform application of the evaluative criteria. In each instance the relationship appeared to be linear, with proportions of positive responses decreasing from the Upper Third to the Lower Third of participating districts.

When procedures were considered, there was high agreement across all levels that the principal served as the primary evaluator of teachers. More diversity was indicated when participants were asked who should serve as the primary evaluator, but differences were not significant. Significant relationships were found across district levels when teachers and administrators were asked about the frequency of formal evaluation. Even though both groups reflected this relationship, the direction of it varied across positions, with teachers at the Lower Third reporting more frequent observations and administrators at the Upper and Middle Thirds reporting more frequent observations. Also, the Lower Third's teachers reported the lowest per cent of short observations while the Upper Third's administrators reported the highest percentage of long observations. Differences in these cases were not significant when comparisons were made across district levels, however. All groups of teachers

agreed that whatever the amount of time spent in observation of teachers, it was adequate. When administrators were considered separately, those of the Middle Third indicated that the sample of behavior was not adequate for making judgments. The use of peer and self-evaluation also was significantly related to district levels, and in the same apparent direction as noted above when job targets and teacher involvement in the selection of criteria were discussed (increasing from Lower to Upper Third). Student ratings were seldom reported. There was no significant difference regarding the use of inservice time to help teachers improve in specific areas of weakness. In general, all districts reported little time spent in this regard. Finally, it was noted that there was a significant relationship across district levels concerning perceived teacher involvement in the determination of evaluation procedures; the Lower Third indicated limited teacher involvement.

While the small size of the samples from the individual school districts did not allow looking for agreement within districts, it was possible to study the levels of districts for internal consistency. There was general agreement within district levels that the primary purpose of teacher evaluation is and should be the improvement of instruction. There was also agreement that the primary criteria used in evaluation are classroom activities, with the least emphasis on product measures. When differences were noted across positions or grade level (as reported later), the same differences were found within the district levels, but these differences were often not significant. For example, there were glaring differences between teachers and administrators within district levels regarding the existence of personal bias in evaluations

or the possibility that a person involved in an extra-curricular activity might be evaluated on the success of the activity. Disagreements also were indicated across grade levels of teachers when they were asked if the same criteria should be applied to all teachers. There were no significant differences within districts regarding the frequency and length of observations, but a wide range of responses was usually encountered.

Overall, when an attempt was made to determine satisfaction with teacher evaluation, no significant differences were found. One notable exception occurred in the Upper Third when there was a significant difference between grade levels of teachers regarding their satisfaction with current evaluation procedures.

In attempting to determine whether there was general agreement between teachers and administrators, several interesting results were noted. While there was no significant difference between these groups when selecting the most and least important categories of criteria, it was apparent when the responses to individual criteria were noted that administrators were more unified on their choices of classroom activities, with many process criteria receiving at least ninety per cent of the administrators' support. There was no significant difference between teachers and evaluators as to whether teachers have been involved in the selection of evaluative criteria, but there was a significance when when subjects were asked whether teachers had been involved in the determination of evaluation procedures.

Specific procedures produced a number of differences across positions. The existence of personal bias and evaluation of extra-curricular activities

rather than teaching were both denied by administrators but supported by a large enough percentage of teachers that highly significant differences were found. Teachers and administrators also differed over the frequency and length of observations although the questions did not allow direct comparisons. Another significant difference was found when administrators indicated that procedures were updated regularly and a substantial number of teachers did not agree. There was agreement, however, that the main purpose of teacher evaluation is and should be improvement of instruction. Contradictory results were noted when subjects were asked about their satisfaction with current evaluation practices. Administrators indicated that they were less satisfied with evaluation procedures than teachers, but teachers selected negative symbols to represent evaluation significantly more often.

When teachers were compared across grade levels, there was general agreement on most items. Some exceptions were noted. Teachers differed across grade levels when asked whether the same criteria should be applied to all teachers, with secondary teachers dissenting more frequently. There was also disagreement over who should serve as the evaluator. Secondary teachers indicated more options, but differences were not found to be significant. A significant difference between grade levels was found in the Upper Third when teachers were asked to indicate their satisfaction with evaluation procedures, but this difference was not reflected over all teachers.

Satisfaction with evaluation practices is not easily determined. When subjects were asked directly whether they were satisfied with evaluation procedures, 63.2 per cent of the teachers and 58.3 per cent of the

administrators indicated that they were, thus leaving more than 35 per cent of the teachers and 41 per cent of the administrators not satisfied. Similarly, when asked if the sample of behavior taken during typical observation was adequate, about one-third of the administrators and over 45 per cent of the teachers recorded negative responses. Finally, when subjects selected symbols to represent teacher evaluation, over 20 per cent selected negative ones, with a significant difference found between teachers and administrators.

Conclusions

The data from this study have revealed that there was significant differences across district levels, across positions, and across grade levels of teachers in several areas of teacher evaluation practices. Some of these differences were not unexpected since administrators and teachers have different perspectives from which to view teacher evaluation. However, the frequency of differences across district levels has indicated that the McNally criteria do discriminate between districts on a large number of variables, and that the evaluation materials supplied by the districts are a good source for determining many features of an evaluation program. It would also appear that the evaluative criteria which participants selected tend to follow conventional wisdom, but do not necessarily reflect the findings of research. Finally, it can be noted that if the true purpose of evaluation is the improvement of instruction, this should be reflected in the priority given to helping teachers improve. The responses especially regarding inservice time, did not reveal this to be the case.

Recommendations

Even though the "ideal" evaluation program, even by McNally's standards, may not exist, much can be learned by considering those districts whose programs were highly ranked. Some distinguishing elements of these districts seemed to be increased teacher involvement in determining evaluation practices, and the use of peer and self-evaluation in addition to administrative ratings. These districts also appear to make more frequent use of job targets, an approach which seems best able to lead to instructional improvement while providing for the individual differences of teachers.

One measurement problem which seems to transcend district boundaries is that of obtaining a truly adequate sample of teaching behavior. Since staff time is limited, it would seem that frequent annual observations by administrators are out of the question. Administration ratings could be augmented through the use of a variety of evaluation sources including students and peers. The best compromise is to perform formal evaluation on a two or three year cycle, making multiple administrator observations at that time, rather than to perform single observations over short time spans on an annual basis. Annual evaluation with a focus on improvement can then be carried out using any available source of information. Job targets can help focus these interim evaluations on areas where individual teachers need specific improvement.

In conclusion, it would seem that many of the problems discovered between teachers and administrators can be corrected in time by having a school commitment to evaluation for the improvement of instruction and then by improving communication between teachers and administrators so

that all know exactly what evaluations are supposed to be accomplishing. This requires total staff involvement, and an atmosphere of mutual trust and respect all coupled with a desire to improve.

Recommendations for Future Research

For districts who wish to improve their current evaluation programs, it is recommended that they perform a system-wide self-analysis using an instrument similar to the one employed in this study. However, to encourage respondents to give honest replies, it is necessary to remove the threat which arises when teachers feel that they will be singled out if their replies are contrary to what the school administration wishes to find. This can be done by using the services of an outside researcher to carry out the survey and to combine data into aggregate form, thus insuring the anonymity of individual teachers. While performing this research, the author found possible disclosure to be a very real fear of some participants, and one which had to be allayed before subjects would take part in the study.

On a broader scale, it would be interesting to determine whether the criteria which are claimed are the same as the criteria which administrators actually use. Also, methods of evaluation could be studied to determine ways to improve staff communication and staff satisfaction with evaluation programs.

BIBLIOGRAPHY

Abell, Allen. 1968. The uses and abuses of correlational and regression techniques in the evaluation and prediction of teacher effectiveness. Pages 48-57 in A.S. Barr, D.A. Worchester, Allan Abell, Clarence Beecher, Leland E. Jensen, Archie L. Peronto, Thomas A. Ringness, and John Schmid, Jr. Wisconsin studies of the measurement and prediction of teacher effectiveness. 4th printing. Dembar Publications, Inc., Madison, Wisconsin.

Airisian, Peter W. 1973. Performance-based teacher education: Evaluation issues. Pages 12-20 in Performance education-assessment. The University of the State of New York. The State Education Department, Division of Teacher Education and Certification and Multi-State Consortium on Performance-Based Teacher Education.

Aleamoni, Lawrence M. 1973. Evaluation by students to identify general instructional problems. Office of instructional resources, Illinois University, Urbana.

Anderson, Scarvin B., Samuel Bell, Richardt T. Murphy, and associates. 1975. Encyclopedia of educational evaluation. Jossey-Bass Publishers, San Francisco, California.

Armstrong, Harold R. 1973. Performance evaluation. The National Elementary Principal 52(5):51-55.

Bain, Richard J., Ronald L. Billings, and W. Robert Houston. 1973. Houston needs assessment system. Needs assessment results, teacher competency analysis, Dallas Independent School District. Texas Educational Renewal Center, Houston University, Texas. ERIC ED 091 382.

Baker, Jack, and Gail Langor Karwoski. 1974. The MSA (Minnesota Student Assoc.) method of teacher evaluation. Twin Cities Student Assembly, Minneapolis, Minn. ERIC ED 091 988.

Ballard, Michael, J. Reardon, and Linda Nelson. 1976. Student and peer rating of faculty. Teaching of Psychology 3(2):88-20.

Baltus, Dale F. 1974. Accountable evaluation for improvement. Paper presented at annual convention of Nat. Assoc. of Elem. Prin. ERIC ED 901 859.

Barr, A.S. 1968a. The nature of the problem. Pages 5-9 in A.S. Barr, D.A. Worcester, Allan Abell, Clarence Beecher, Leland E. Jensen, Archie L. Peronto, Thomas A. Ringness, and John Schimd, Jr. Wisconsin studies of the measurement and prediction of teacher effectiveness. 4th printing. Dembar Publications, Inc., Madison, Wisconsin.

Barr, A.S. 1968b. The personal prerequisites to teacher effectiveness. Pages 99-106 in A.S. Barr, D.A. Worcester, Allan Abell, Clarence Beecher, Leland E. Jensen, Archie L. Peronto, Thomas A. Ringness, and John Schmid, Jr. Wisconsin studies of the measurement and prediction of teacher effectiveness. 4th printing. Dembar Publications, Inc., Madison, Wisconsin.

Barr, A.S. 1968c. Teacher effectiveness and its correlates. Pages 134-152 in A.S. Barr, D.A. Worcester, Allan Abell, Clarence Beecher, Leland E. Jensen, Archie L. Peronto, Thomas A. Ringness, and John Schmid, Jr. Wisconsin studies of the measurement and prediction of teacher effectiveness. 4th printing. Dembar Publications, Inc., Madison, Wisconsin.

Barr, A.S., D.A. Worcester, Allan Abell, Clarence Beecher, Leland E. Jensen, Archie L. Peronto, Thomas A. Ringness, and John Schmid, Jr. 1968. Wisconsin studies of the measurement and prediction of teacher effectiveness. 4th printing. Dembar Publications, Inc., Madison, Wisconsin.

Berliner, David C. 1975. Impediments to the study of teacher effectiveness. Far West Laboratory for Educational Research and Development, San Francisco, Cal. ERIC ED 128 343.

Biddle, Bruce J. 1964. The integration of teacher effectiveness research. Pages 2-40 in Bruce J. Biddle and William J. Ellena, eds. Contemporary research on teacher effectiveness. Holt, Rinehart and Winston, New York.

Biddle, Bruce J. and William J. Ellena, eds. 1964. Contemporary research on teacher effectiveness. Holt, Rinehart and Winston, New York.

Blalock, Herbert M., Jr. 1972. Social statistics. 2nd ed. McGraw-Hill Book Co., New York.

Bodine, Richard. 1973. Teacher's self-assessment. Pages 169-173 in Ernest R. House, ed. School evaluation: The politics & process. McCutchan Publishing Corp., Chicago.

Bohlken, Robert and Kim Giffin. 1971. A paradigm for determining high school teacher effectiveness. Paper presented at 56th annual meeting of Speech Communications Assoc., New Orleans, La. ERIC ED 050 118.

Bolen, John E. 1973a. The dilemma in evaluating instruction. The Educational Digest 38(1):20-22.

Bolen, John E. 1973b. The dilemma in evaluation instruction. The National Elementary Principal 52(5):72-75.

Bolton, Dale L. 1973a. Collecting evaluation data. The National Elementary Principal 52(5):77-86.

Bolton, Dale L. 1973b. Selection and evaluation of teachers. McCutchan Publishing Corp., Berkeley, Cal.

Bradley, Ruth, and others. 1964. Measuring teacher competence. Research background and current practices. California Teachers Assoc., Burlingame, Cal. ERIC ED 040 152, 1970.

Brophy, Jere E. 1974. Achievement Correlates. Pages 33-46 in Herbert J. Walberg, ed. Evaluating educational performance: A sourcebook of methods, instruments, and examples. McCutchan Publishing Corp., Berkeley, Cal.

Broudy, H.S. 1969. Can we define good teaching? The Record-Teachers College 7-(7):583-592.

Brown, Ric. 1977. The relationship between student evaluation of teaching, student achievement, and student perceptions of teacher effectiveness. ERIC ED 133 314.

Bryson, Rebecca. 1974. Teacher evaluation and student learning: A reexamination. The Journal of Educational Research 68(1):12-14.

Burdin, Joel. 1974. Three views of competency based teacher education: I theory. Gastback no. 48. The Phi Delta Kappa Educational Foundation, Bloomington, Indiana.

Burnett, J. Dale. 1975. Evaluating teachers. Research and information report. Paper presented at annual meeting of the Canadian Society for the Study of Education. University of Alberta, Edmonton, Alberta. ERIC ED 114 934, 1976.

Campbell, Lloyd P., and John A. Williamson. 1974. Teaching competencies for supervising teachers. North Texas State Univ., Denton, Texas. ERIC ED 100 935.

Carlson, Robert V. and Richard Park. 1976. Teacher evaluation: Relevant concepts and related procedures. ERIC ED 129 739.

Centra, John A. 1972. What instructors learn from students: Student and faculty ratings of instruction. Educational Testing Service, Princeton, N.J. ERIC ED 065 033.

Chisholm, Mary G. 1977. Provocative opinion: Student evaluation: The red herring of the decade. Journal of Chemical Education 54(1):22-23.

Coats, William D. and Lloyd Swierenga. 1972. Student perceptions of teachers - a factor analytic study. The Journal of Educational Research 65(8):357-360.

Cohen, Arthur M. and Florence B. Brewer. 1969. Measuring faculty performance. American Assoc. of Junior Colleges, Washington, D.C. ERIC ED 031 222.

Coker, Homer. 1976. A competency based teacher certification system. ERIC ED 123 247.

Coker, Homer, and Joan G. Coker. 1975. A competency based certification system. School of Education, West Georgia College, Carrollton. ERIC ED 128 342, 1977.

Commission of Public School Personnel Policies in Ohio. 1972. Teacher evaluation to improve learning. The fourth report. Commission on Public School Personnel policies in Ohio, Cleveland, Ohio.

Coney, Robert C. 1972. Performance evaluation. Thrust for Educational Leadership 2(1):11-13.

Cook, J. Marvin. 1972. Direct measurement of collegiate teaching teaching effectiveness. Educational Technology 12(6):51-54.

Cornwell, C.D. 1974. Statistical treatment of data from student teaching evaluation questionnaires. Journal of Chemical Education 51(3):115-160.

Cox, Richard C. 1974. Confusion between norm-referenced and criterion-referenced measurement. Phil Delta Kappan 55(5):319+.

Crim, Roger D. 1974. The use of the video recorder in teacher self-evaluation. ERIC ED 106 270.

Davis, Hazel. 1964. Evolution of current practices in evaluating teacher competence. Pages 41-66 in Bruce J. Biddle and William J. Ellena, eds. Contemporary research on teacher effectiveness. Holt, Rinehart and Winston, New York.

Davis, Sam, and Barbara Gross-Davis. 1974. Assessing teacher effectiveness based on student learning gains. California Journal of Educational Research 25(4):186-191.

DeVaughn, J. Everette. 1971. Policies, procedures and instruments in evaluation of teachers and administrator performance. Georgia State Department of Education, Robert Davis Associates, Inc., Atlanta, Ga. ERIC ED 059 151.

Doyle, Kenneth O., and Susan E. Whitely. 1974. Student ratings as criteria for effective teaching. American Educational Research Journal 11(3):259-274.

Dressel, Paul L. 1976. Handbook of academic evaluation. Jossey-Bass Publishers, San Francisco, Cal.

Dropkin, Stan. 1973. Twin dimensions of teacher behavior assessment. Contemporary Education 44(3):142-146.

Drumheller, Sidney J. 1974. Evaluating teachers through a jaundiced eye. Educational Technology 14(7):17-22.

Drummond, William H. Involving the teacher in evaluation. The National Elementary Principal 52(5):30-32.

Dunkin, Michael, and Bruce Biddle. 1974. Findings for process variables. Pages 362-410 in The study of teaching. Holt, Rinehart and Winston, Inc., New York.

Dwyer, Francis M. 1973. Selected criteria for evaluating teacher effectiveness. Improving College Teaching 21(1):51-52.

Eads, Albert Edward, Jr. 1974. A study of the attitudes of teachers toward a supervision of objectives teacher evaluation model. Ph.D. Dissertation. University of South Carolina. 125 pp. (Dissertation Abstracts International 36(02):0628-A).

Eastridge, Harry E. 1976. Student evaluation and teacher performance. NASSP (Nat. Assoc. of Sec. Sch. Prin.) Bulletin 60(401):48-54.

Egnatoff, John G. ca. 1974. A new look at evaluating teacher performance. In Teacher Effectiveness Package TE-3. Institute for Educational Services, Inc., Bedford, Mass.

Ellett, Lowell E., and Earl P. Smith. 1975. Improving performance of classroom teachers through videotaping and self-evaluation. AV Communication Review 23(3):277-288.

Elliott, Velma L. 1974. Peer evaluation for teachers? Why not? Elementary English 51(5):727-730.

Ellman, Niel. 1976. Evaluating representative teaching behavior. NAASSP (Nat. Assoc. of Sec. Sch. Prin.) Bulletin 40(401):25-27.

ERIC Abstracts. 1973. A collection of ERIC document resumes on competency-based evaluation of educational staff. ERIC Abstracts Series No. 28. ERIC ED 084 651.

Fahey, George L. 1970. Student rating of teaching - some questionable assumptions. Pages 8-20 in Student evaluation of teaching. Presentations at a conference. Institute for Higher Education, Pittsburgh University, Pa. ERIC ED 054 724.

Fast, Elizabeth T. 1974. In-service staff development as a logical part of performance evaluation. School Media Quarterly 3(1):35-41.

Feldhusen, John F. 1976. A model of instruction as the base for course and instructor evaluation. College Student Journal 10(3):197-203.

Feldman, Sandra. 1972. Teacher evaluation - A teacher unionist's view. ETS (Ed. Testing Service), Princeton, N.J. ERIC ED 086 647.

Flanders, Ned A. 1964. Some relationships among teacher influence, pupil attitudes, and achievement. Pages 196-231 in Bruce J. Biddle and William J. Ellena, eds. Contemporary research on teacher effectiveness. Holt, Rinehart and Winston, New York.

Flanders, Ned A. 1970. Analyzing teaching behavior. Addison-Wesley Publishing Co., Reading, Mass.

Flanders, Ned A. 1974. The changing base of performance-based teaching. Phi Delta Kappan 55(5):312-315.

Flanders, Ned A., and Greta Morine. 1975. The assessment of proper control and suitable learning environment. Pages 67-100 in Madan Mohan and Ronald E. Hull. Teacher effectiveness: Its meaning, assessment, and improvement. Educational Technology Publications, Englewood Cliffs, N.J.

Fortune, Jimmie C., Dennis E. Hinkle, Robert B. Frary, and Janice K. McBee. 1975. Use of classroom distributions of student achievement test scores to evaluate the instructional effectiveness of teachers. ERIC ED 117 170, 1976.

Fox, Willard, and Ronald D. Jones. 1970. A plan for evaluation of teaching efficiency through cooperative goal-setting. The Clearing House 44(9):541-543.

Freund, John E. 1967. Modern elementary statistics. 3rd ed. Prentice-Hall, Inc., Englewood Cliffs, N.J.

Frey, Peter W. 1973. Student ratings of teaching: Validity of several rating factors. Science 182(4107):83-84.

Gage, N.L., ed. 1963. Handbook of research on teaching. Rand McNally and Co., Chicago.

Gage, N.L. 1972. Teacher effectiveness and teacher education: The search for a scientific basis. Pacific Books, Publishers, Palo Alto, Cal.

Gaines, W. George. 1973. What focus in the measurement of teaching competency? ERIC ED 092 571.

Glass, Gene V. 1974. Teacher effectiveness. Pages 11-32 in Herbert J. Walberg, ed. Evaluating educational performance: A sourcebook of methods, instruments, and examples. McCutchan Publishing Corp., Berkeley, Cal.

Gonzaga University, School of Ed. 1973. Personalized performance-based teacher education program. Spokane, Washington. ERIC ED 085 385.

Gray, Frank. 1975. How successful is performance evaluation? Paper presented at 107th Annual Convention of the American Assoc. of School Administrators, Dallas, Texas.

Haak, Ruth Adolf, and others. 1972. Student evaluation of teacher instrument, II manual. Research and Development Center for Teacher Education, Texas University, Austin. ERIC ED 080 574.

Halbert, Susan J. 1975. Student evaluation of teachers and administrators. Oregon School Student Council Bulletin 18(8):1-3.

Hall, George L., Jr. 1974. Assessing staff effectiveness: Practical approaches to meaningful evaluation. ERIC ED 097 338.

Hamachek, Don. 1975a. Characteristics of good teachers and implications for teacher education. Pages 239-251 in Madan Mohan and Ronald E. Hull. Teacher effectiveness: Its meaning, assessment, and improvement. Educational Technology Publications, Englewood Cliffs, N.J.

Hamachek, Don E. 1975b. Personal characteristics of good teachers. Pages 293-327 in Don E. Hamachek. Behavior dynamics in teaching, learning, and growth. Allyn and Bacon, Inc., Boston.

Hammons, Abe L. 1975. How to evaluate teachers on performance. Paper presented at the 35th annual meeting of the National School Boards Assoc., Miami Beach, Fla.

Hanke, John E. 1973. The art of teacher evaluation. Improving College and University Teaching 21(1):56-57.

Hastings, J. Thomas. 1973. Evaluation of instruction. Pages 140-145 in Ernest R. House, ed. School evaluation: The politics & process. McCutchan Publishing Corp., Berkeley, Cal.

Hayman, John L. Jr., and Rodney N. Napier. 1975. Evaluation in the schools: A human process for renewal. Brooks/Cole Publishing Co., Monterey, Cal.

Herman, Jerry J. 1973. Developing an effective school staff evaluation program. Parker Publishing Co., Inc., West Nyack, N.Y.

Hidlebaugh, Everett. ca. 1973. Teacher performance evaluation items. Mimeographed. Author, Dike, Iowa.

Hildebrand, Milton. 1972. How to recommend promotion for a mediocre teacher without actually lying. Journal of Higher Education 43(1): 44-62.

Hogan, Thomas P. 1973. Similarity of student ratings across instructors, courses, and time. Research in Higher Education 1(2):148-154.

Horton, Margaret W. 1977. The use of teacher evaluation to improve instruction. Paper presented at annual meeting of the Assoc. of Teacher Education, Atlanta, Ga. ERIC ED 135 775.

House, Ernest R., ed. 1973. School evaluation: The politics & process. McCutchan Publishing Corp., Berkeley, Cal.

House, Peggy. 1975. Balancing the equation on CBTE. The Mathematics Teacher 68(7):519-524.

Houstan, W. Robert, and Robert B. Howsam. 1974. CBTE: The ayes of Texas. Phi Delta Kappan 55(5):294-313.

Houstan, W. Robert, and Howard L. Jones. 1974. Three views of competency-based teacher education: II University of Houstan. Fastback No. 49. Phi Delta Kappa Educational Foundation, Bloomington, Indiana.

Howsam, Robert B. 1973. Current issues in evaluation. The National Elementary Principal 52(5):12-17.

Hoyt, Donald P. 1969. Identifying effective teaching behaviors. Kansas State University, Manhattan, Kansas. ERIC ED 039 197, 1970.

Hoyt, Donald P. 1973. Measurement of instructional effectiveness. Research in Higher Education 1(4):367-378.

Jenkins, Joseph R., and R. Barker Bausell. ca. 1972. How teachers view the effective teacher: Student learning is not the top criterion. Reprinted. Source unknown.

Johnson, Henry C., Jr. 1976. Court, craft, and competence: A reexamination of "teacher evaluation" procedures. Phi Delta Kappan 57(9): 606-611.

Johnson, James A., and Byron F. Radebaugh. 1969. Excellent teachers - What makes them outstanding? The Clearing House 44(3):152-156.

Jones, Anthony S. 1972. A realistic approach to teacher evaluation. The Clearing House 46(8):474-481.

Kelley, Edgar A. 1974. Three views of competency-based teacher education: III University of Nebraska. Fastback No. 50. The Phi Delta Kappa Foundation, Bloomington, Indiana.

Kenny, James, Gertrude Hentshel, and Kathy Elpers. 1972. How students see teachers. ERIC ED 077 921.

Kirk, Roger E. 1968. Experimental design: Procedures for the behavioral sciences. Brooks/Cole Publishing Co., Belmont, Cal.

Klein, Susan S. 1974. What is R and D personnel competency assessment? Paper presented at the annual meeting of the Amer. Educational Research Assoc., Chicago, Ill. ERIC ED 095 203.

Krasno, Richard M. 1972. Accountability and research on teacher effectiveness. Administrator's Notebook 21(1):1-4.

Kulik, James A., and Chen-Lin C. Kulik. 1974. Student ratings of instruction. Teaching of Psychology 1(2):51-57.

Lancaster, Otis E. 1973. Measuring teacher effectiveness. IEEE Transactions on Education EI6(3):138-142.

Laurits, James. 1967. Thoughts on the evaluation of teaching. Pages 32-42 in Pi Lambda Theta. The evaluation of teaching. Geo. Banta Co., printer, Washington, D.C.

Levinson, Harry. 1970. Management by whose objectives? Harvard Business Review 48(4):125-134.

Lewis, James, Jr. 1973. Appraising teacher performance. Parker Publishing Co., West Nyack, N.Y.

Lucio, William H. 1973. Pupil achievement as an index of teacher performance. Educational Leadership 31(1):71-77.

Mallery, David. 1975. The strengths of a good school faculty: Notes on evaluation, growth, and professional partnership of teachers. National Assoc. of Independent Schools, Boston, Mass. ERIC ED 102 719.

Manatt, Richard P., Kenneth L. Palmer, and Everett Hidlebaugh. 1976. Evaluating teacher performance with improved rating scales. NASSP (Nat. Assoc. of Sec. School Prin.) 60(401):21-24.

Marten, Gary Stewart. 1975. Teacher and administrator attitudes toward evaluation and systematic classroom observation. Ph.D. dissertation. University of Oregon. 162 pp. (Dissertation Abstracts International 36:5780-A).

Matthews, Marvin R. 1972. Educational accountability to whom - for what? Thrust for Educational Leadership 2(1):5-10.

Maxwell, W. David. 1974. PBTE: A case of the emperor's new clothes. Phi Delta Kappan 55(5):306-311.

McDonald, Frederick J. 1972. Evaluation of teaching behaviors. In W. Robert Houstan and Robert B. Howsam, eds. Competency-based teacher education progress, problems, and prospects. Science Research Associates, Inc., Chicago.

McDonald, Frederick J. 1973. The state of the art in performance assessment of teaching competence. Pages 21-27 in Performance education-assessment. The State Education Department, Division of Teacher Education and Certification and Multi-State Consortium on Performance-Based Teacher Education, The University of the State of New York.

McDonald, Frederick J. 1974. The national commission on performance-based education. Phi Delta Kappan 55(5):296-298.

McGuire, Christine. 1973. A proposed model for the evaluation of teaching. Pages 85-108 in Pi Lambda Theta. The evaluation of teaching. Geo. Banta Co., Printer, Washington, D.C.

McKeachie, W.J. 1970. Research on student ratings of teaching. Pages 1-7 in Student evaluation of teaching. Presentations at a conference. Institute for Higher Education, Pittsburgh University, Pa. ERIC ED 054 724.

McKeachie, W.J. 1971. Studies of student ratings of faculty. Final report. College of Lit., Sci., and Arts, Office of Ed., Michigan University, Ann Arbor. ERIC ED 057 745.

McKeachie, W.J., Yi-Guang Lin, and William Mann. 1971. Student ratings of teacher effectiveness: Validity studies. American Educational Research Journal 8(3):435-444.

McKenna, Bernard H. 1973a. A context for teacher evaluation. The National Elementary Principal 52(5):18-23.

McKenna, Bernard H. 1973b. Teacher evaluation - some implications. Today's Education 62(2):55-56.

McNally, Harold J. 1973. What makes a good evaluation program. The National Elementary Principal 52(5):24-29.

McNeil, John D. 1967. Concomitants of using behavioral objectives in the assessment of teacher effectiveness. The Journal of Experimental Education 36(1):69-74.

McNeil, John D. 1971a. Performance tests: Assessing teachers of reading. Paper presented to annual meeting of the Cal. Educ. Res. Assoc., San Diego, Cal. ERIC ED 054 200.

McNeil, John D. 1971b. Toward accountable teachers - their appraisal and improvement. Holt, Rinehart and Winston, Inc., New York.

Medley, Donald M. 1963. Experiences with the OScAR technique. The Journal of Teacher Education 14(3):267-273.

Medley, Donald M. 1971. Indicators and measures of teaching effectiveness: A review of the research. ERIC ED 088 844.

Medley, Donald M. 1973. A process approach to teacher evaluation. The National Elementary Principal 52(5):33-35.

Medley, Donald M. and Harold E. Mitzel. 1963. Measuring classroom behavior by systematic observation. Pages 247-328 in N.L. Gage, ed. Handbook of research on teaching. Rand McNally and Co., Chicago.

Medley, Donald M., and others. 1975. Assessment and research in teacher education: Focus on PBTE. PBTE monograph series no. 17. Amer. Assoc. of Colleges for Teacher Ed., Washington, D.C. ERIC ED 107 600.

Meux, Milton and B. Othanel Smith. 1964. Logical dimensions of teaching behavior. Pages 127-164 in Bruce J. Biddle and William J. Ellena, eds. Contemporary research on teacher effectiveness. Holt, Rinehart and Winston, New York.

Meyer, Herbert H., Emanuel Kay, and John R.P. French, Jr. 1965. Split roles in performance appraisal. Harvard Business Review 43(1): 123-129.

Miller, Clint and Dorothy Miller. 1971. The importance of certain personal qualities and professional competencies in successful classroom teaching. The Journal of Teacher Education 22(1):37-39.

Miller, Richard I. 1972. Evaluating faculty performance. Jossey-Bass, Inc., Publishers, San Francisco, Cal.

Millman, Jason. 1974. Criterion referenced measurement. Pages 309-397 in W. James Popham, ed. Evaluation in education: Current applications. McCutchan Publishing Corp., Berkeley, Cal.

Mocker, Donald W. 1974. A report on the identification, classification, and ranking of competencies appropriate for adult basic education teachers. School of Ed., Missouri University, Kansas City. ERIC ED 099 469.

Mohan, Madan and Ronald E. Hull. 1975. Teaching effectiveness: Its meaning, assessment, and improvement. Educational Technology Publications, Englewood Cliffs, N.J.

Mood, Alexander M. 1970. Do teachers make a difference? Paper presented at a conference sponsored by the Off. of Ed., Bureau of Ed. Professions Development. ERIC ED 040 253.

Musella, Donald. 1970. Improving teacher evaluation. The Journal of Teacher Education 21(1):15-21.

Naperville Community Unit School District no. 203. 1973. Teacher performance evaluation handbook. Board of education. Dupage and Will Counties, Illinois.

National Center for Educational Communication. 1971. Teacher Evaluation PREP-21. Office of Education (DHEW), Washington, D.C. ERIC ED 044 546.

National School Boards Association, Educational Policies Services. 1971. School board policies on teacher evaluation educational policies development kit. ERIC ED 058 657.

Neel, Thomas E. 1972. Classroom performance standards. Thrust for Educational Leadership 2(1):17-20.

New Mexico State Department of Education. 1973. Competency based certification, January 1, 1973. Action in New Mexico. Interim report. Sante Fe. ERIC ED 099 367.

Nie, Norman H., C. Hadlar Hull, Jean G. Jenkins, Karin Steinbrenner, and Dale H. Bent. 1975. SPSS statistical package for the social sciences. 2nd ed. McGraw-Hill Book Company, New York.

Nutt, Hubert Wilbur. 1920. The supervision of instruction. Houghton Mifflin Co., Boston.

O'Connor, Edward F., Jr. and Thomas B. Justiz. 1970. Identifying the effective instructor. California University, Los Angeles. ERIC 035 416.

O'Hanlon, James and Lynn L. Morteson. 1977. Improving teacher evaluation. The CEDR Quarterly 10(4):3-7.

Peck, Robert F. and Donald J. Veldman. 1973. Personal characteristics associated with effective teaching. Research and Development Center for Teacher Education, Texas University, Austin. ERIC ED 078 038.

Pedersen, K. George. 1975. Improving teacher effectiveness. Education Canada 15(3):12-20.

Penn-Harris-Madison School Corporation. 1973. Performance-based teacher appraisal program for the Penn-Harris-Madison School Corporation. Author, Mishawaki, Ind. ERIC ED 093 919.

Pennsylvania State Department of Education. 1974. Pennsylvania's approach to the competency certification of educational professional staff. Working paper. Harrisburg. ERIC ED 099 373.

Peronto, Archie L. 1968. The abilities and patterns of behaviors of good and poor teachers. Pages 88-98 in A.S. Barr, D.A. Worcester, Allen Abell, Clarence Beecher, Leland E. Jensen, Archie L. Peronto, Thomas A. Ringness, and John Schmid, Jr. Wisconsin studies of the measurement and prediction of teacher effectiveness. 4th printing. Dembar Publications, Inc., Madison, Wis.

Pierce, Wendell and Ronald Smith. 1974. The race to evaluate. Instructor 83(6):5-12.

Pi Lambda Theta. 1967. The evaluation of teaching. Geo. Banta Co., Printer, Washington, D.C.

Pine, Gerald J. and Angelo V. Boy. 1975. Necessary conditions for evaluating teachers. NASSP (Nat. Assoc. of Sec. School Prin.) Bulletin 59(395):18-23.

Pitman, John C. 1974. A resource guide; national and regional developments in competency based teacher education (CBTE) and competency based certification (CBC). Rhode Island State Department of Education, Providence, R.I. ERIC ED 110 454.

Pittsburgh University, Pa., Institute for Higher Education. 1970. Student evaluation of teaching. Presentations at a conference. ERIC ED 054 724.

Place, Roger A. 1974. Removing the incompetent practitioner. Paper presented at Amer. Assoc. of School Admin. annual convention, 106th. Atlantic City, N.J. ERIC ED 088 237.

Pohlman, John T. 1975. A description of teaching effectiveness as measured by student ratings. Journal of Educational Measurement 12(1):49-54.

Poliakoff, Lorraine L. 1973a. Evaluating school personnel today. ERIC Clearinghouse on Teacher Education, Washington, D.C. ERIC ED 073 045.

Poliakoff, Lorraine. 1973b. Recent trends in evaluating school personnel. The National Elementary Principal 52(5):39-44.

Popham, W. James. 1971a. Designing teacher evaluation systems. The Instructional Objectives Exchange, Los Angeles, Cal.

Popham, W. James. 1971b. Teaching skill under scrutiny. Phi Delta Kappan 52(10):597-612.

Popham, W. James. 1972. California's precedent-setting teacher evaluation law. Educational Research 1(7):13-15.

Popham, W. James. 1973a. Alternative teacher assessment strategies. Pages 34-38 in The University of the State of New York. Performance education-assessment. The State Education Department, Division of Teacher Education and Certification and Multi-State Consortium on Performance-Based Teacher Education.

Popham, W. James. 1973b. Identification and assessment of minimal competencies for objectives-oriented teacher education programs. Pages 51-57 in The University of the State of New York. Performance education-assessment. The State Education Department, Division of Teacher Education and Certification and Multi-State Consortium on Performance-Based Teacher Education.

Popham, W. James. 1973c. Teaching performance tests. The National Elementary Principal 52(5):56-59.

Popham, W. James, ed. 1974a. Evaluation in education: Current applications. McCutchan Publishing Corp., Berkeley, Cal.

Popham, W. James. 1974b. Teacher evaluation and domain-referenced measurement. Educational Technology 14(6):33-37.

Popham, W. James. 1975a. Applications of teaching performance tests in pre-service and inservice teacher education. Journal of Teacher Education 26(3):244-248.

Popham, W. James. 1975b. Educational evaluation. Prentice Hall Inc., Englewood Cliffs, N.J.

Popham, W. James and Eva L. Baker. 1966. A performance test of teaching effectiveness. ERIC ED 014 449, 1968.

Powell, J.P. 1970. Teaching and interacting. Contemporary Education 41(3):137-138.

Quirk, Thomas J. 1974. Some measurement issues in competency-based teacher education. Phi Delta Kappan 55(5):316-319.

Raisch, C. Daniel. 1973. Competency based teacher evaluation: A study and model. Ph.D. Dissertation. Miami University, Fla. 221 pp. (Dissertation Abstracts International 35:0948-A).

Reavis, Ralph G. 1975. Teacher evaluation practices in selected Texas public high schools. Ph.D. Dissertation. North Texas State University. 128 pp. (Dissertation Abstracts International 36:7978-A).

Redfern, George B. 1972. How to evaluate teaching - a performance objectives approach. School Management Institute, Worthington, Ohio.

Redfern, George B. 1973. Legally mandated evaluation. The National Elementary Principal 52(5):45-50.

Remmers, H.H. 1963. Rating methods in research on teaching. Pages 329-378 in N.L. Gage, ed. Handbook of Research on Teaching. Rand McNally and Co., Chicago.

Resnick, Norman H. and Henry R. Reinert. 1973. Competency based teacher evaluation. Colorado Journal of Educational Research 13(1):19-21.

Ritenour, Louise R., ed. 1975. Students' opinions help improve college instruction. ETS (Educational Testing Service) Developments 22(3).

Ritenour, Louise R., ed. 1977. ETS researchers find teachers do make a difference in student achievement. ETS (Educational Testing Service) Developments 24(1).

Rodin, Miriam. 1973. Research - Can students evaluate good teaching? Change 5(6):66-67, 80.

Rodin, Miriam and Burton Rodin. 1972. Student evaluation of teachers. Science 177(9):1164-1166.

Rosencranz, Howard A. and Bruce J. Biddle. 1964. The role approach to teacher competence. Pages 232-263 in Bruce J. Biddle and William J. Ellena, eds. Contemporary research on teacher effectiveness. Holt, Rinehart and Winston, New York.

Rosenshine, Barak. 1970. The stability of teacher effects upon student achievement. Review of Educational Research 40(5):647-662.

Rosenshine, Barak. 1973. PBTE: Proceed with caution. Pages 28-33 in Performance education-assessment. The State Education Department, Division of Teacher Education and Certification and Multi-State Consortium on Performance-Based Teacher Education, The University of the State of New York.

Rosenshine, Barak. 1975. Enthusiastic teaching: A research review. Pages 105-120 in Madan Mohan and Ronald E. Hull. Teacher effectiveness: Its meanings, assessment, and improvement. Educational Technology Publications, Englewood Cliffs, N.J.

Rosenshine, Barak and Norma Furst. 1975. Research on teacher performance criteria. Pages 37-72 in B.O. Smith, ed. Research in teacher education: A symposium. Prentice-Hall Inc., Englewood Cliffs, N.J.

Rosenshine, Barak and Marilyn Marten. 1974. Teacher education and teaching behavior: Comments on the state-of-the-research. Educational Researcher 3(7):11-12.

Rosenshine, Barak and Barry McCaw. 1973. Assessing teachers in public education. Pages 146-155 in Ernest R. House, ed. School evaluation: The politics & process. McCutchan Publishing Corp., Berkeley, Cal.

Rosner, Benjamin and Patricia M. Kay. 1974. Will the promise of C/PBTE be fulfilled? Phi Delta Kappan 55(5):290-295.

Roth, Robert A. 1975. The nature of and alternatives for teacher competency statements and implications for assessment techniques. Teacher Preparation and Professional Development. Michigan State Dept. of Education, Lansing. ERIC ED 117 110.

Roth, Robert A. and Peggy Mahoney. 1975. Teacher competencies and assessment techniques. Paper presented at the annual meeting of the Amer. Educational Research Assoc., Washington, D.C. ERIC ED 104 838.

Ryans, David G. 1960. Characteristics of teachers. American Council on Education, Washington, D.C.

Ryans, David G. 1964. Research on teacher behaviors in the context of the teacher characteristics study. Pages 67-101 in Bruce J. Biddle and William J. Ellena, eds. Contemporary research on teacher effectiveness. Holt, Rinehart and Winston, New York.

Ryans, David G. 1967. Teacher behavior can be evaluated. Pages 43-64 in Pi Lambda Theta. The evaluation of teaching. Geo. Banta Co., Printer, Washington, D.C.

Saadeh, Ibrahim Q. 1970. Teacher effectiveness or classroom efficiency: A new direction in the evaluation of teaching. The Journal of Teacher Education 21(1):73-89.

Saif, Philip S. 1974. Teacher evaluation. Capital Region Education Council, West Hartford, Conn. Bureau of Elementary and Secondary Ed., Office of Ed. (DHEW), Washington, D.C. ERIC ED 100 974.

Saif, Philip. 1976. Teacher evaluation. Connecticut Assoc. of Boards of Education, Inc., Hartford, Conn. ERIC ED 126 090.

Sandefur, J.T. and Alex A. Bressler. 1971. Classroom observation systems in preparing school personnel. In J.T. Sandefur and Alex A. Bressler. Interaction analysis: Selected papers. Washington Assoc. of Teacher Educators in collaboration with ERIC Clearinghouse on Teacher Education.

Sandoz, Ellis. 1974. CBTE: The nays of Texas. Phi Delta Kappan 55(5): 304-305.

Sarthory, Jose A. 1973. Professional improvement and staff evaluation. An information paper on KSA 72-9001 to 72-9006: Evaluation of certified school employees. Kansas State Dept. of Education, Topeka. ERIC ED 089 402.

Schalock, H.D., Jesse H. Carrison, and Bert Y. Kersh. 1973. From commitment to practice in assessing the outcome of teaching: A case study. In Performance education-assessment. The State Education Dept., Division of Teacher Education and Certification and Multi-State Consortium on Performance-Based Teacher Education, The University of the State of New York.

Schmid, John. 1968. Factor analysis of the teaching complex. Pages 58-69 in A.S. Barr, D.A. Worcester, Allan Abell, Clarence Beecher, Leland E. Jensen, Archie L. Peronto, Thomas A. Ringness, and John Schmid, Jr. Wisconsin studies of the measurement and prediction of teacher effectiveness. 4th printing. Dembar Publications, Inc., Madison, Wis.

Scott, Craig S. and Gaylord Thorne, Comps. 1974. Assessing faculty performance: A partially annotated bibliography. Oregon State System of High Education, Monmouth, Teaching Research Division. ERIC ED 093 187.

Seldon, David. 1969. Evaluate teachers? QuEST papers series no. 4. American Federation of Teachers, Washington, D.C.

Shea, Neil Patrick. 1977. A study of teacher evaluation process in the junior high schools of the twenty largest school districts in Kansas. Ed.D. Thesis. University of Kansas. 87 pp. (Dissertation Abstracts International 38:3868-A).

Simon, Anita and Gil Boyer, eds. 1967-70. Mirrors for behavior: An anthology of classroom observation instruments. Vol. I-XIV. Research for Better Schools, Inc., Philadelphia, Pa.

Simpkins, W.S., R.K. Browne, and T.W. Field. 1973. Teacher differences as perceived by students. Improving College and University Teaching 21(1):64-66.

Sinatra, William James. 1975. An investigation into the relationship between teacher evaluation and the interpersonal compatibility of the teacher and the evaluator. Ph.D. Dissertation. State University of New York at Buffalo. 173 pp. (Dissertation Abstracts International 36:6415-A).

Slobin, Dan Y. and David G. Nichols. 1969. Student rating of teaching. Improving College and University Teaching 17(4):244-248.

Smith, B. Othanel. 1967. Teaching: Conditions of its evaluation. Pages 65-84 in Pi Lambda Theta. The evaluation of teaching. Geo. Banta Co., Printer, Washington, D.C.

Smith, Owen T. 1972. Case study No. 3. Evaluating the teacher. Thrust for Educational Leadership 2(1):30-31.

Soar, Robert S. and Ruth M. Soar. 1975. Problems in using pupil outcomes for teacher evaluation. National Education Assoc., Washington, D.C.

Sorenson, Garth and Cecily F. Gross. 1967. Teacher appraisal, a matching process. California University, Los Angeles. Center for the Study of Evaluation of Instruction Program. ERIC ED 016 299, 1968.

Spady, William G. and Douglas Mitchell. 1977. Competency based education: Organizational issues and implications. Educational Researcher 6(2): 9-15.

Speicher, Dean. 1972. Can teacher evaluation be made meaningful? Paper presented at Amer. Assoc. of School Administrators annual meeting, Atlantic City, N.J. ERIC ED 060 575.

Spencer, Ralph L. and William E. Boyd. 1977. CBTE is succeeding in the State of New York. Phi Delta Kappan 58(9):677+.

Spillane, Robert R. and Dorothy Levenson. 1976. Teacher training: A question of control, not content. Phi Delta Kappan 57(7):435-439.

Spivey, James R. 1976. Evaluation by objectives - a model for teacher evaluation. NASSP (Nat. Assoc. of Sec. School Prin.) Bulletin 60(401): 40-44.

Thomas, Donald. 1974. The principal and teacher evaluation. NASSP (Nat. Assoc. of Sec. School Prin.) Bulletin 58(386):1-7.

Thornsley, Jerome R. 1972. Recognition and respect for teacher competence. Thrust for Educational Leadership 2(2):23-27.

Tolor, Alexander. 1973. Evaluation of perceived teacher effectiveness. Journal of Educational Psychology 64(1):98-104.

Tomblin, Elizabeth Aline Risinger. 1976. Effects of participatory and non-participatory methods of teacher evaluation on selected teacher variables. Ph.D. Dissertation. University of Colorado, Boulder. 143 pp. (Dissertation Abstracts International 37:4771-A).

Tuckman, Bruce W. and Wilmot F. Oliver. 1968. Effectiveness of feedback to teachers as a function of source. Journal of Educational Psychology 50(4):297-301.

Tye, Kenneth A. 1972. Performance evaluation: Another link in the self-fulfilling prophecy? Thrust for Educational Leadership 2(1):21-23.

The University of the State of New York. 1973. Performance education-assessment. The State Education Department, Division of Teacher Education and Certification and Multi-State Consortium on Performance-Based Teacher Education, Author.

Vail, Robert B. 1974. AACTE leadership training institute on performance-based teacher education. ERIC ED 099 355.

Vogt, Karl E. and Harry Lasher. 1973. Does student evaluation stimulate improved teaching? College of Business Administration, Bowling Green State University, Ohio. ERIC ED 078 748.

Walberg, Herbert J., ed. 1974. Evaluating educational performance: A sourcebook of methods, instruments, and examples. McCutchan Publishing Corp., Berkeley, Cal.

Washington, Eva. 1970. The expert teacher action study: A new approach to teacher evaluation. The Journal of Teacher Education 21(2):258-263.

Washington, Eva. 1971. The expert teacher program. Fearon Publishers, Belmont, Cal.

Watman, Thomas J. 1972. Supervision for growth. The Clearing House 46(9):567-568.

Webber, Wilhelmina Sweet. 1976. An analysis of teachers' perceptions of teacher evaluation practices in Palm Beach County, Florida. Ed.D. Thesis. Florida Atlantic Univ. 107 pp. (Dissertation Abstracts International 37:4773-A).

Weisenstein, Greg R. 1976. Teacher evaluation: The principal's role. OSSC (Oregon School Study Council) Bulletin 20(3). ERIC ED 130 438.

Wicks, Larry E. 1973. Opinions differ: Teacher evaluation. Today's Education 62(3):42-43.

Wilson, Laval S. 1974. Assessing teacher skills: Necessary component of individualization. Phi Delta Kappan 56(3):207-209.

Wilson, Laval S. 1975. How to evaluate teacher performance. Paper presented to annual convention of National School Boards Assoc., 35th, Miami Beach, Fla.

Wilson, Robert G. 1972. Teaching effectiveness: Its measurement. Journal of Engineering Education 62(6):550-552.

Wolf, Robert L. 1973. How teachers feel toward evaluation. Pages 156-168 in Ernest R. House, ed. School evaluation: The politics & process. McCutchan Publishing Corp., Berkeley, Cal.

Wood, Randy. ca. 1974. A use of the Q-sort technique in educational evaluation. ERIC ED 128 360.

Worcester, D.A. 1968. Some assumptions, explicitly and implicitly made in the investigations here summarized. Pages 120-133 in A.S. Barr, D.A. Worcester, Allan Abell, Clarence Beecher, Leland E. Jensen, Archie L. Peronto, Thomas A. Ringness, and John Schmid, Jr. Wisconsin studies of the measurement and prediction of teacher effectiveness. 4th printing. Dembar Publications, Inc., Madison, Wis.

Zax, Manuel. 1971. Outstanding teachers: Who are they? The Clearing House 45(5):285-289.

Zelby, Leon W. 1974. Student-faculty evaluation. Science 183(20):1267-1270.

Zelenak, Melchoir James. 1973. Teacher perception of the teacher evaluation process. Ph.D. Dissertation. The University of Iowa, Iowa City. 116 pp. (Dissertation Abstracts International 34:2944-A).

ACKNOWLEDGMENTS

It is fitting at this point to mention some of the people who have helped make this work a possibility. First of all, I would like to thank my wife, Lois, for the time she spent in proofreading and correcting the original manuscript as well as for the support which she has provided. I would also like to express my thanks to those members of the Iowa State faculty, as well as to several Mathematics Department graduate students. I would like to especially thank Dr. Rex Thomas for his aid with the computer, his contribution to Question 40, and his statistical insight. Finally, I would like to express my deepest appreciation to the man who, without knowing it, led me into this inquiry. Mr. Cal Callaway, an assistant principal at Oregon High School, Oregon, Wisconsin, showed me by his determination to be a better evaluator that there are ways to perform teacher evaluation which can make it a rewarding rather than a threatening experience for teachers. His commitment to this end, in spite of other duties, has served as a prime motivation.

APPENDIX A. SELECTION OF PARTICIPATING SCHOOL DISTRICTS BY APPLYING THE MCNALLY CRITERIA

Ranking of Schools by McNally Criteria

McNally Criteria / Schools	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. Purposes written	1	1	½	1	1	0	0	0	1	½	1	1	1	0	1	1	1	1	1	1	1
2. Policies reflect research	1	0	0	1	0	0	0	0	0	0	0	½	1	0	0	1	½	1	1	0	1
3. Teachers know criteria	1	1	1	1	1	1	1	0	1	½	1	1	1	0	1	1	0	1	1	1	1
4. Evaluation done cooperatively	1	0	½	½	0	0	0	0	0	0	0	1	1	0	0	1	0	0	1	0	1
5. Valid and reliable																					
a. important to learner	½	½	½	½	½	½	½	½	1	½	½	½	½	0	½	½	0	½	0	½	½
b. adequate sample	1	0	0	0	0	0	0	½	0	0	0	0	½	0	½	1	0	0	0	0	1
c. evaluators agree	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
d. guidelines	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	1	0	1	½	0	1
e. limits on criteria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
6. More diagnostic than judgmental	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	1	½	1	1	0	1
7. Self-evaluation	1	0	0	1	0	0	0	0	0	0	0	½	1	0	0	1	0	0	1	0	1
8. Self-image enhanced	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1
9. Encourages creativity	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	1
10. Personalized feedback	1	1	1	0	1	½	½	0	0	½	0	½	1	0	0	1	½	1	½	½	1
11. Part of inservice	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
School totals	8½	4½	2½	6	5½	2	2	1	3	2	2½	5½	11	0	3	10½	2½	8½	10	3	12½
Schools selected for the sample	X	X				X		X			X*	X	X		(X)*				X		X

*School number 15 requested to be omitted from the study and was replaced by school number 11.

Districts within levels: Upper Third - 13, 19, 21 Middle Third - 1, 2, 12 Lower Third - 6, 8, 11

APPENDIX B. FIRST COVER LETTER, QUESTIONNAIRE, FOLLOW-UP LETTER

Dear Colleague:

You have been selected to be a participant in a survey of nine Iowa school districts on the topic of teacher evaluation. Your name was chosen from a list of teachers and administrators supplied by your district. In order to minimize the efforts of any one school district, only thirty teachers and six administrators were chosen per district. Thus, your honest and complete responses are needed if this is to be a representative study.

The study itself is an attempt to determine exactly what is being done today in the area of teacher evaluation and how this evaluation is being carried forward. You will be asked questions dealing with the procedures used in your building, the purposes of teacher evaluation in your school, the kinds of criteria which are used, and your satisfaction or lack of satisfaction with your evaluation system. Again, I would stress, I am particularly concerned with your opinions.

It should be noted that you are under no obligation to complete this questionnaire. However, I would emphasize that precautions have been taken to insure that the information gathered by this survey will be kept confidential. Your individual reply will be organized in aggregate form with others from your grade level and school. No information from an individual's response will be released. No information which would use the name of a specific school will be printed without additional written consent from the school district.

When you complete your questionnaire, put the instrument and the scan sheet into the envelope and return it sealed to the district office through the school mail. You should remove your name tag to insure confidentiality.

If there are any questions regarding this survey, you may contact me at my office, 445 Carver Hall, Iowa State University, Ames, Iowa, 50011, or by phone at the office (515) 294-8184 or at home (after 5:30 p.m.) (515) 232-3692. I plan to visit your school district during the first few days of May to collect the completed questionnaires from the district office. I would be most happy to visit with you at that time.

Thank you for your time. It is my hope that the information which we obtain by this survey will help us improve the evaluation programs of many school districts. Your help is deeply appreciated.

Sincerely,

William K. Tomhave

William K. Tomhave
Educational Researcher

A. J. Netusil wkt

A. J. Netusil
Major Professor

Teacher Evaluation Questionnaire

The following information is to be used for the purposes of statistical comparison and validation of our sampling. Mark the letter of the appropriate response in the space provided.

____ 1. Your present position: A. Classroom Teacher B. Building Administrator.

____ 2. Your grade level: A. Elementary (K-6) B. Secondary (7-12).

____ 3. Years of experience within this system (including this year): A. 1-2 B. 3-5 C. 6-10 D. 11-15 E. over 15.

____ 4. Years of experience- total (including this year): A. 1-2 B. 3-5 C. 6-10 D. 11-15 E. over 15.

This next section deals with evaluative criteria. In order to simplify the recording of the data, you are asked to mark your responses on the scan sheet provided. You will note that your sheet has already been coded with an I.D. number. This number is used solely to keep your responses together, and will not be used to identify your responses from the completed survey.

Below you will find a list of statements taken from a wide variety of evaluation instruments as well as from the research literature. These statements are generally listed under a heading of "criteria," "evidences of effective teaching," or some similar heading. You are asked to consider each statement and respond to the question: "Is this a standard by which teachers are evaluated in my school?" You are not being asked whether these things are things which teachers do, nor are you being asked whether these things should be used as standards. Rather, you are being asked whether these statements are currently used as standards for judging a teacher's effectiveness within your school. The criteria which you select need not be part of your school's written criteria, but must be among the standards actually used by evaluators in your building.

On the computer sheet provided please mark your response to each statement as follows:

Mark A if you know that the statement is definitely used as a standard.

Mark B if you feel that the statement is probably used as a standard.

Mark C if you feel that the statement is probably not used as a standard.

Mark D if you know that the statement is definitely not used as a standard.

Mark E if you have no opinion.

You may use any lead pencil which will mark darkly. Do not use pen as the machine used to read the sheets will read pen marks as blanks.

You will note that the criteria are divided into five categories, and that some of the categories have many more statements than others. This is not meant to indicate the importance of the category in so far as this survey is concerned, but rather reflects the relative abundance of statements in that particular category.

I. Personal Characteristics:

1. The teacher presents a physical appearance appropriate for the teaching assignment.
2. The teacher practices cleanliness and good grooming and adopts suitable dress and manner.
3. The teacher is physically and emotionally able to perform required duties.
4. The teacher has regular attendance.
5. The teacher has a positive, enthusiastic attitude.
6. The teacher meets classroom situations with a sense of humor.
7. The teacher demonstrates consistency and reliability as well as flexibility and adaptability.
8. The teacher is free from irritating habits.
9. The teacher is self-confident.
10. The teacher is punctual.
11. The teacher is reliable.
12. The teacher has no distracting mannerisms.
13. The teacher is tactful.
14. The teacher possesses a business-like or task oriented behavior.
15. The teacher's handwriting is clear and legible.
16. The teacher maintains neatness of desk, materials, boards, and files.

II. Professional Qualities: "Is this a standard by which teachers are evaluated in my school?"

17. The teacher belongs to organizations and attends meetings in his/her professional area.
18. The teacher takes applicable professional advancement courses.
19. The teacher reads professional literature.
20. The teacher attends meetings of school related organizations.
21. The teacher continues improvement through research and experimentation.
22. The teacher participates in grade level and departmental meetings.
23. The teacher participates in other district instructional meetings.
24. The teacher undertakes new projects- i.e. is willing to be innovative.
25. The teacher demonstrates a positive attitude toward teaching.
26. The teacher promotes the positive value of education.
27. The teacher assesses each lesson and unit in terms of student responses to techniques, activities, and materials, and in terms of student achievement of objectives.
28. The teacher is involved in self-assessment for the improvement of instruction.
29. The teacher accepts responsibilities such as hall duty, extra-curricular duties, service on committees, etc.
30. The teacher promotes good student citizenship.
31. The teacher cooperates with the local parent-teacher organization.
32. The teacher maintains the confidences of students and staff.
33. The teacher complies with the rules and regulations of the school system.
34. The teacher follows established communication channels.
35. The teacher provides constructive criticism to the administration and accepts the same from them.
36. The teacher takes responsibility for accurately communicating the school program to the community.
37. The teacher meets the workday time requirement.
38. The teacher submits required reports promptly and accurately.
39. The teacher respects school property.
40. The teacher accepts suggestions from specialists and is willing to try them.
41. The teacher is a good team worker.
42. The teacher is critical of and constantly trying to improve his/her work.
43. The teacher assumes responsibilities outside the classroom as they relate to school.
44. The teacher uses discretion in discussing school affairs.
45. The teacher works to improve the school image.
46. The teacher is readily available to students.
47. The teacher is up-to-the-minute on current developments and teaching methods.
48. The teacher adapts new findings and techniques for use in his/her classes.
49. The teacher frequently volunteers to help.

III. Classroom Activities:

50. The teacher has written objectives.
51. The teacher prepares sufficient and appropriate written lesson plans for himself/herself or a substitute.
52. The teacher develops plans consistent with the short and long range goals and objectives of the course.
53. The teacher uses appropriate activities and resources to meet the needs of varied student or group abilities and interests.
54. The teacher arranges the classroom appropriately for class activities.
55. The teacher has needed materials and equipment ready for use.
56. The teacher attempts to provide a well ventilated, clearly lighted classroom which contributes to a positive learning atmosphere.
57. The teacher provides for individual differences.
58. The teacher is aware of individual differences.
59. The teacher uses non-contact time for class preparation.
60. The teacher makes appropriate revisions in content to include up-to-date information.
61. The teacher uses objectives to determine instructional activities and evaluations.
62. The teacher exhibits a mastery of subject matter appropriate for the grade or skill level taught
63. The teacher knows and uses content which is appropriate to the subject area.
64. The teacher is familiar with and uses community resources to enrich the educational program.
65. The teacher uses content appropriate to the students' abilities and needs.
66. The teacher states his/her goal expectations clearly.
67. The teacher develops lessons and makes assignments which are clear and consistent with course goals.
68. The teacher makes clear and concise explanations.
69. The teacher gives clear directions.
70. The teacher relates current lessons to previous learning.
71. The teacher displays and encourages creativity.
72. The teacher motivates pupils to advance at their own optimum rates.
73. The teacher emphasizes the importance of applying acquired skills and knowledge.
74. The teacher uses illustrations from contemporary life.
75. The teacher involves students.

III. Classroom Activities (continued): "Is this a standard by which teachers are evaluated in my school?"

76. The teacher makes different assignments based on diagnostic information.
77. The teacher individualizes instruction: this includes differentiated assignments, alternative methods of learning, and varied expectations.
78. The teacher uses praise and positive reinforcement.
79. The teacher uses his/her voice effectively.
80. The teacher provides for large group instruction, small group instruction, and independent study.
81. The teacher uses instructional media.
82. The teacher provides opportunities for students to direct some of their own learning.
83. The teacher provides freedom for students to move about.
84. The teacher is well organized but flexible.
85. The teacher maintains an efficient, orderly routine.
86. The teacher exhibits a friendly but positive control of the class.
87. The teacher is firm, consistent, and fair.
88. The teacher moves among students during the working period when appropriate.
89. The teacher communicates class rules and regulations.
90. The teacher supports building and district discipline policies.
91. The teacher inspires a spirit of cooperation among students.
92. The teacher provides for student participation in planning behavior standards.
93. The teacher provides for continual supervision of students.
94. The teacher provides an atmosphere where students remain at task.
95. The teacher maintains a continuous record of student progress.
96. The teacher uses tests that reflect course objectives.
97. The teacher uses tests that provide for the variety of abilities in the class.
98. The teacher uses a clear, reasonable, and fair grading system.
99. The teacher provides for individual differences by diagnosing pupil needs.
100. The teacher employs the participation of students in evaluation of instructional practices and individual growth.
101. The teacher makes continual, daily assessment and observation of students.
102. The teacher uses the results of classroom tests to improve classroom instruction.
103. The teacher makes the classwork interesting- puts his/her material across in an interesting way.
104. The teacher conducts a classroom in which pupils actively participate in classroom discussion and activities.
105. The teacher uses a wide variety of materials to supplement the basic program.
106. The teacher evaluates pupils in terms of their academic, social and emotional growth.
107. The teacher gives class objectives that are reflected in the exams he/she gives.
108. The teacher adequately explains answers to exam questions after an exam has been given.
109. The teacher has an excellent subject matter background and uses initiative to stay ahead in the field.
110. The teacher encourages active participation and recognizes the instructional value of his/her actions.

IV. Interpersonal Relations:

111. The teacher shows respect for and interest in all pupils as individuals.
112. The teacher uses constructive criticism and is supportive of students.
113. The teacher allows students to make constructive criticism.
114. The teacher is available to students and offers additional assistance.
115. The teacher is fair, impartial, and objective in his/her treatment of students.
116. The teacher makes provision for pupil participation in both planning and evaluation when appropriate.
117. The teacher allows time for students to share worthwhile ideas and experience.
118. The teacher sees that there is a feeling of good will in the class.
119. The teacher is understanding, encouraging, and helpful.
120. The teacher provides security for students.
121. The teacher makes each child feel important and respected.
122. The teacher has the respect of the students.
123. The teacher has a good rapport with the students.
124. The teacher actively listens to students.
125. The teacher calls each student by name.
126. The teacher avoids sarcasm.
127. The teacher strives to make strdents aware of their progress throughout the reporting period.
128. The teacher uses discretion and respect when speaking of colleagues.
129. The teacher cooperates with the administration and keeps them informed of pertinent situations.
130. The teacner utilizes and works cooperatively and courteously with para-professionals and other support staff.
131. The teacher accepts constructive criticism and guidance.

IV. Interpersonal Relations (continued): "Is this a standard by which teachers are evaluated in my school?"

132. The teacher shares ideas and materials willingly with other staff members.
133. The teacher shows a genuine respect and concern for colleagues.
134. The teacher complies with authorized policies and procedures.
135. The teacher is willing to accept his share of building responsibilities.
136. The teacher accepts group decisions and abides by them.
137. The teacher deals effectively with parents.
138. The teacher keeps parents informed of each student's progress and problems in school.
139. The teacher works cooperatively with parents for optimal student growth.
140. The teacher actively works for positive school- community relations.

V. Product Measures:

141. The teacher is responsible for class achievement of course objectives.
142. The teacher is responsible for students' achievement in later courses.
143. The teacher is responsible for students' success in later life.
144. The teacher is responsible for students' attitudes toward his/her class.
145. The teacher is responsible for students' attitudes toward his/her subject.
146. The teacher is responsible for students' attitudes toward school in general.
147. The teacher is responsible for students' attitudes toward the community.
148. The teacher is responsible for students' attitudes toward life.
149. The teacher is responsible for the attitudes of parents toward the school.
150. The teacher is responsible for the community's attitude toward the school.
151. The teacher is responsible for the ability of his/her students to find employment.
152. The teacher is responsible for giving students means of finding satisfaction in life.
153. The teacher is responsible for developing social responsibility in pupils.
154. The teacher is responsible for developing good citizenship in pupils.
155. The teacher is responsible for the success of students in extra-curricular competition (i.e. music, speech, drama, athletics, etc.).

This completes the material which is to be recorded on the scan sheet. The remainder of the questions will be recorded by the number of the question as you did for numbers 1 through 4.

Within each category of criteria given above, select the three <u>most important</u> criteria, ranked in order of importance from most to least, by listing the criterion number next to the number of the category.

____ 5. Personal Characteristics	____ 6. Professional Qualities	____ 7. Classroom Activities
____	____	____
____	____	____
____ 8. Interpersonal Relations	____ 9. Product Measures	
____	____	
____	____	

____ 10. Of the five categories of statements listed above, which area do you feel is most frequently used as the primary source of evaluative data in your school?
A. Personal Characteristics B. Professional Qualities C. Classroom Activities
D. Interpersonal Relations E. Product Measures

____ 11. Of the five categories listed in number 10, which area do you feel is considered the <u>least</u> in making evaluative decisions?

____ 12. Have the criteria which are used to evaluate teachers in your school been carefully explained to the teachers? A. Yes B. No

____ 13. Were teachers involved in the selection of evaluative criteria? A. Yes B. No

____ 14. Are the criteria written? A. Yes B. No

____ 15. Are administrators supplied with guidelines for the uniform application of the criteria during evaluation? A. Yes B. No

____ 16. Are the same criteria (standards) used for all teachers? A. Yes B. No

____ 17. Should the same criteria be used to evaluate all teachers? A. Yes B. No

____ 18. Are the criteria used in your evaluations based upon "Job targets" or some kind of managemen by objectives approach? A. Yes B. No

____ 19. Should teachers be involved in determining what criteria will be used in the evaluation of their teaching? A. Yes B. No

____ 20. Are teacher evaluations in your building relatively free from personal biases of the evaluator? A. Yes B. No

____ 21. Is a person who is involved with a popular extra-curricular activity (athletics, music, drama, etc.) more likely to be evaluated on the success of that activity than on his/her teaching? A. Yes B. No

---

The following questions deal with the procedures which are used within your system to evaluate teachers. Answer each question to the best of your knowledge, and write the letter of the appropriate response in the space provided.

____ 22. From this list of possible teacher evaluators, select the letter of the person(s) who serve as the primary evaluator in your building:
A. Principal/Assistant Principal B. Supervisor/Curriculum Specialist C. Department Head
D. Peers (fellow teachers) E. Students F. Central Office Personnel
G. Others (specify) _____.

____ 23. Of those personnel listed in number 22, who do <u>you</u> think is best equiped to evaluate teachers?

If you are a teacher:

____ 24t. How many times have you been formally observed this past year?
A. 0 B. 1 C. 2 D. 3-4 E. 5 or more
If your answer was <u>A</u>, indicate when you were last observed: _____

____ 25t. During your last formal observation, how long did the observer spend in you classroom?
A. Less than 15 minutes B. Between 15 minutes and 30 minutes C. Between 30 minutes and 45 minutes D. Between 45 and 60 minutes E. Over 60 minutes

____ 26t. Do you feel that this was adequate to make a judgement regarding your ability as a teacher?
A. Yes B. No

If you are an administrator:

____ 24a. How frequently do you formally observe the typical teacher in your building?
A. Three or more times a year B. Twice a year C. Annually D. Semi-Annually
E. Every three years or less

____ 25a. How long is your average observation? A. Less than 15 minutes B. Between 15 and 30 minutes
C. Between 30 and 45 minutes D. Between 45 and 60 minutes E. Over 60 minutes

____ 26a. Do you feel that this is adequate for making sound judgments? A. Yes B. No

____ 27. Below are listed a variety of means for collecting evaluation data. Write the letter which precedes the method most frequently used in your building.
A. Administrative rating forms B. Student rating forms C. Systematic Observation (i.e. Interaction Analysis) D. Self-evaluation E. Video or Audio tape F. Student Achievement data G. Peer ratings H. Other (specify)_____.

In Teacher Evaluation:

____ 28. Is peer evaluation used? A. Frequently B. Sometimes C. Never

____ 29. Is self-evaluation used? A. Frequently B. Sometimes C. Never

____ 30. Are students' ratings used? A. Frequently B. Sometimes C. Never

In Teacher Evaluation:

____ 31. Is student achievement used? A. Frequently B. Sometimes C. Never.

____ 32. Virtually every evaluation program mentions "informal observation." How great a role should this play in the total evaluation of a teacher?
A. Very important B. Important C. Relatively unimportant D. Insignificant.

____ 33t. If you are a teacher, in your opinion does your evaluator have sufficient training to be an effective evaluator? A. Yes B. No C. Don't know.

____ 33a. If you are an administrator, how much formal training do you have which you would say specifically prepared you to evaluate teachers?
A. Less than 5 hours B. 5 to 8 hours C. 9 to 16 hours D. 16 to 40 hours E. More than 40 hrs.

____ 34. How many hours of in-service training are devoted each school year to help teachers improve in specific areas of weakness as shown by their evaluations?
A. 0-1 B. 2-4 C. 5-8 D. 9-16 E. Over 16.

____ 35. Have teachers been involved in determining what procedures will be used to evaluate their teaching? A. Yes B. No.

____ 36. Is the evaluation procedure itself evaluated and improved regularly? A. Yes B. No.

____ 37. Are you satisfied with the evaluation procedures used in your building? A. Yes B. No.

Below is a list of some of the possible purposes which can be served by teacher evaluation. By number 38 write the letter which corresponds to the purpose which you see as the main purpose of teacher evaluation as practiced in your school. By number 39 write the letter of the purpose which you feel teacher evaluation ought to serve in your school.

A. Renewal/non-renewal B. Staff assignment C. Staff reduction D. Improvement of instruction
E. Differentiated pay F. Other (specify) _____.

____ 38. Current purpose.

____ 39. Desired purpose.

____ 40. If you were asked to select a symbol to represent teacher evaluation, what would it be?
A. A yard stick B. A guillotine C. A stethoscope D. A Peeping Tom
E. A final examination F. A psychiatrist's couch G. Numbers in a hat
H. A popularity pageant.

Dear Colleague:

Recently you received a copy of a questionnaire on the topic of teacher evaluation. At this time I have not received your response so I am sending this second instrument. If you have completed the first questionnaire and have forwarded it to the district office, please ignor this second survey. If you have completed the questionnaire but have not as yet sent it to the district office, please send it directly to me in the addressed, stamped envelope provided.

If you chose to not complete the first questionnaire due to its length, you can still benefit the study by answering only the forty questions which do not require the computer scan sheet. Simply answer those questions directly on the instrument and mail it to me.

If you still prefer to not participate, then merely indicate that on the questionnaire and return the papers to me in the envelope.

I am looking forward to your response. To date the number of replies has been very encouraging.

REMEMBER: YOUR OPINION DOES MATTER !!

Thank you for your time.

Sincerely,

William K. Tomhave

William K. Tomhave
Educational Researcher

APPENDIX C. EVALUATIVE CRITERIA - SURVEY RESPONSE

Evaluative Criteria

Below is a list of evaluative criteria as they appeared on the survey instrument. After each criterion number are listed three rows of information. The first row contains the per cent of the total replies which selected a given response. The second and third display the same information for teachers and administrators respectively.

	A	B	C	D	E	tot.
1.	44.6	37.7	12.1	4.8	0.9	231
	47.6	37.7	11.5	3.7	1.0	191
	30.0	45.0	15.0	10.0	0.0	40
2.	37.7	45.5	10.4	5.2	1.3	231
	37.7	46.6	9.9	4.2	1.6	191
	37.5	40.0	12.5	10.0	0.0	40
3.	56.0	37.5	3.9	1.7	0.9	232
	55.2	39.1	4.2	1.0	0.5	192
	60.0	30.0	2.5	5.0	2.5	40
4.	35.3	43.1	14.2	4.3	3.0	232
	32.8	46.4	15.1	3.1	2.6	192
	47.5	27.5	10.0	10.0	5.0	40
5.	58.4	31.2	8.7	1.7	0.0	231
	57.1	33.0	8.4	1.6	0.0	191
	65.0	22.5	10.0	2.5	0.0	40
6.	17.7	48.9	22.9	5.6	4.8	231
	15.2	46.6	27.2	5.8	5.2	191
	30.0	60.0	2.5	5.0	2.5	40
7.	54.5	37.2	4.3	3.0	0.9	231
	52.9	38.7	4.7	3.1	0.5	191
	62.5	30.0	2.5	2.5	2.5	40
8.	6.1	39.4	31.6	10.4	12.6	231
	6.3	37.7	33.5	8.4	14.1	191
	5.0	47.5	22.5	20.0	5.0	40
9.	39.8	45.9	10.8	2.6	0.9	231
	38.2	45.5	12.6	2.6	1.0	191
	47.5	47.5	2.5	2.5	0.0	40

	A	B	C	D	E	tot.
22.	45.7	40.5	12.1	1.7	0.0	232
	47.4	38.5	13.0	1.0	0.0	192
	37.5	50.0	7.5	5.0	0.0	40
23.	26.7	31.9	27.2	10.8	3.4	232
	25.0	30.7	29.2	11.5	3.6	192
	35.0	37.5	17.5	7.5	2.5	40
24.	33.2	42.7	15.9	6.9	1.3	232
	32.8	43.2	15.6	6.8	1.6	192
	35.0	40.0	17.5	7.5	0.0	40
25.	62.9	29.3	4.7	2.2	0.9	232
	62.5	30.7	3.6	2.1	1.0	192
	65.0	22.5	10.0	2.5	0.0	40
26.	44.8	39.7	10.3	2.6	2.6	232
	43.8	40.6	9.9	2.6	3.1	192
	50.0	35.0	12.5	2.5	0.0	40
27.	34.9	37.6	17.0	8.3	2.2	229
	33.9	37.0	19.0	7.4	2.6	189
	40.0	40.0	7.5	12.5	0.0	40
28.	31.1	39.5	20.6	6.1	2.6	228
	31.7	39.5	21.2	5.3	2.6	189
	28.2	41.0	17.9	10.3	2.6	39
29.	53.7	34.6	6.5	4.3	0.9	231
	50.8	36.6	7.9	3.7	1.0	191
	67.5	25.0	0.0	7.5	0.0	40
30.	37.1	39.3	17.0	5.7	0.9	229
	33.3	41.8	18.0	5.8	1.1	189
	55.0	27.5	12.5	5.0	0.0	40
31.	17.7	29.7	26.3	15.5	10.8	232
	17.7	30.7	24.0	15.6	12.0	192
	17.5	25.0	37.5	15.0	5.0	40
32.	35.5	47.4	10.5	4.8	1.8	228
	34.4	48.7	11.1	3.7	2.1	189
	41.0	41.0	7.7	10.3	0.0	39
33.	65.4	28.4	5.2	1.3	0.0	231
	65.6	28.6	5.2	0.5	0.0	192
	64.1	25.6	5.1	5.1	0.0	39

	A	B	C	D	E	tot.
10.	41.1	40.7	13.9	3.9	0.4	231
	41.7	40.1	14.6	3.1	0.5	192
	38.5	43.6	10.3	7.7	0.0	39
11.	56.9	36.2	5.2	1.3	0.4	232
	56.3	37.0	5.7	0.5	0.5	192
	60.0	32.5	2.5	5.0	0.0	40
12.	3.9	40.9	34.5	11.6	9.1	232
	3.6	38.0	38.0	10.4	9.9	192
	5.0	55.0	17.5	17.5	5.0	40
13.	19.4	53.4	20.3	5.6	1.3	232
	19.8	52.1	20.8	5.7	1.6	192
	17.5	60.0	17.5	5.0	0.0	40
14.	34.2	47.6	13.4	2.6	2.2	231
	33.5	46.6	15.2	2.1	2.6	191
	37.5	52.5	5.0	5.0	0.0	40
15.	5.6	18.1	43.1	28.0	5.2	232
	5.7	15.6	43.2	29.7	5.7	192
	5.0	30.0	42.5	20.0	2.5	40
16.	12.1	30.6	30.2	22.4	4.7	232
	12.5	27.6	32.3	22.9	4.7	192
	10.0	45.0	20.0	20.0	5.0	40
17.	22.0	27.6	27.6	20.3	2.6	232
	22.9	27.6	28.1	19.3	2.1	192
	17.5	27.5	25.0	25.0	5.0	40
18.	42.2	31.0	17.7	8.2	0.9	232
	44.8	29.2	18.8	6.3	1.0	192
	30.0	40.0	12.5	17.5	0.0	40
19.	7.8	29.0	30.3	29.0	3.9	231
	8.3	28.6	29.2	29.7	4.2	192
	5.1	30.8	35.9	25.6	2.6	39
20.	19.0	37.7	25.5	15.2	2.6	231
	18.8	36.6	26.2	15.2	3.1	191
	20.0	42.5	32.5	15.0	0.0	40
21.	17.2	27.6	34.9	17.2	3.0	232
	17.7	29.2	33.9	16.1	3.1	192
	13.3	17.8	35.6	20.0	2.2	40

	A	B	C	D	E	tot.
34.	40.8	42.1	10.1	3.9	3.1	228
	40.7	43.4	10.1	3.2	2.6	191
	41.0	35.9	10.3	7.7	5.1	39
35.	22.6	38.7	23.0	10.9	4.8	230
	20.4	37.7	26.2	10.5	5.2	191
	33.3	43.6	7.7	12.8	2.6	39
36.	18.3	31.7	31.7	13.5	4.8	230
	18.3	30.4	33.0	12.6	5.8	191
	17.9	38.5	25.6	17.9	0.0	39
37.	58.4	28.6	8.2	2.6	2.2	231
	58.3	28.6	8.9	2.1	2.1	192
	59.0	28.2	5.1	5.1	2.6	39
38.	42.8	40.2	11.4	3.9	1.7	229
	41.1	41.6	11.6	3.7	2.1	190
	51.3	33.3	10.3	5.1	0.0	39
39.	38.4	37.1	14.4	6.1	3.9	229
	38.4	36.3	15.8	4.7	4.7	190
	38.5	41.0	7.7	12.8	0.0	39
40.	29.6	49.6	11.7	6.1	3.0	230
	28.3	50.3	12.0	5.8	3.7	191
	35.9	46.2	10.3	7.7	0.0	39
41.	40.0	40.4	11.7	5.2	2.6	230
	39.8	41.4	11.5	5.2	2.1	191
	41.0	35.9	12.8	5.1	5.1	39
42.	33.5	45.2	13.9	5.2	2.2	230
	33.5	45.5	14.7	4.2	2.1	191
	33.3	43.6	10.3	10.3	2.6	39
43.	33.2	42.4	13.5	6.1	4.8	229
	32.6	43.7	13.2	5.3	5.3	190
	35.9	35.9	15.4	10.3	2.6	39
44.	27.0	43.0	20.0	6.1	3.9	230
	28.3	42.4	20.4	4.7	4.2	191
	20.5	46.2	17.9	12.8	2.6	39
45.	21.4	45.0	21.0	7.4	5.2	229
	22.6	43.2	22.1	5.8	6.3	190
	15.4	53.8	15.4	15.4	0.0	39

	A	B	C	D	E	tot.
46.	39.7	40.6	14.0	4.4	1.3	229
	40.5	39.5	14.7	3.7	1.6	190
	35.9	46.2	10.3	7.7	0.0	39
47.	18.3	41.7	30.4	5.7	3.9	230
	16.2	42.4	32.5	5.2	3.7	191
	28.2	38.5	20.5	7.7	5.1	39
48.	23.5	39.1	23.9	1.7	1.7	230
	23.0	47.6	25.7	1.6	2.1	191
	25.6	56.4	15.4	2.6	0.0	39
49.	14.8	39.6	30.0	10.4	5.2	230
	15.7	39.3	29.3	9.9	5.8	191
	10.3	41.0	44.4	12.8	2.6	39
50.	43.7	29.7	17.1	8.6	0.9	222
	41.2	27.5	20.9	9.3	1.1	182
	55.0	40.0	0.0	5.0	0.0	40
51.	61.9	22.5	10.0	4.3	1.3	231
	58.1	24.1	11.5	4.7	1.6	191
	80.0	15.0	2.5	2.5	0.0	40
52.	51.1	35.9	8.7	3.0	1.3	231
	49.2	37.2	8.9	3.1	1.6	191
	60.0	30.0	7.5	2.5	0.0	40
53.	59.7	30.7	7.4	1.7	0.4	231
	59.2	30.9	8.4	1.6	0.0	191
	62.5	30.0	2.5	2.5	2.5	40
54.	40.7	37.2	16.9	3.9	1.3	231
	41.4	35.6	18.3	3.7	1.0	191
	37.5	45.0	10.0	5.0	2.5	40
55.	48.1	37.2	10.4	3.0	1.3	231
	40.1	38.7	11.0	2.6	1.6	191
	57.5	30.0	7.5	5.0	0.0	40
56.	30.7	31.6	22.5	10.4	4.8	231
	29.8	30.4	24.1	10.5	5.2	191
	35.0	37.5	15.0	10.0	2.5	40
57.	58.4	32.9	7.4	0.4	0.9	231
	57.1	33.5	8.4	1.0	0.0	191
	65.0	30.0	2.5	2.5	0.0	40

	A	B	C	D	E	tot.
58.	59.3	34.6	4.3	1.3	0.4	231
	59.7	35.1	4.2	0.5	0.5	191
	57.5	32.5	5.0	5.0	0.0	40
59.	19.9	31.6	29.4	13.0	6.1	231
	19.4	30.9	31.4	11.5	6.8	191
	22.5	35.0	20.0	20.0	2.5	40
60.	26.0	45.5	18.6	6.1	3.9	231
	25.1	45.4	18.8	6.3	4.2	191
	30.0	45.0	17.5	5.0	2.5	40
61.	38.5	39.8	13.0	4.3	4.3	231
	37.7	40.8	12.0	4.2	5.2	191
	42.5	35.0	17.5	5.0	0.0	40
62.	65.4	29.0	4.3	0.9	0.4	231
	66.0	28.3	4.7	0.5	0.5	191
	62.5	32.5	2.5	2.5	0.0	40
63.	58.3	33.0	6.1	2.6	0.0	230
	58.1	33.5	6.3	2.1	0.0	191
	59.0	30.8	5.1	5.1	0.0	39
64.	20.0	43.9	27.4	6.1	2.6	230
	20.4	44.5	25.7	6.3	3.1	191
	17.9	41.0	35.9	5.1	0.0	39
65.	56.3	36.4	5.6	1.7	0.0	231
	55.0	37.2	6.3	1.6	0.0	191
	62.5	32.5	2.5	2.5	0.0	40
66.	38.5	36.8	17.7	3.9	3.0	231
	38.7	37.2	16.8	3.7	3.7	191
	37.5	35.0	22.5	5.0	0.0	40
67.	52.4	35.9	8.7	2.6	0.4	231
	51.8	34.6	10.5	2.5	0.5	191
	55.0	42.5	0.0	2.5	0.0	40
68.	48.1	39.8	10.4	1.3	0.4	231
	46.6	40.3	12.0	0.5	0.5	191
	55.0	37.5	2.5	5.0	0.0	40
69.	53.9	34.8	10.0	1.3	0.0	230
	51.3	36.6	12.0	0.0	0.0	191
	66.7	25.6	0.0	7.7	0.0	39

	A	B	C	D	E	tot.
70.	37.8	42.6	14.8	3.0	1.7	230
	35.8	42.6	16.8	2.6	2.1	190
	47.5	42.5	5.0	5.0	0.0	40
71.	33.8	43.7	16.5	3.5	2.6	231
	32.5	44.5	16.8	8.7	2.6	191
	40.0	40.0	15.0	2.5	2.5	40
72.	35.1	40.3	19.5	2.2	3.0	231
	34.6	38.2	22.0	2.1	3.1	191
	37.5	50.0	7.5	2.5	2.5	40
73.	30.3	46.3	16.0	3.9	3.5	231
	29.3	47.6	15.7	3.7	3.7	191
	35.0	40.0	17.5	5.0	2.5	40
74.	16.0	38.4	29.0	10.8	4.8	231
	15.2	39.3	29.8	10.5	5.2	191
	20.0	40.0	25.0	12.5	2.5	40
75.	59.3	35.1	3.5	1.7	0.4	231
	60.2	35.1	3.7	0.5	0.5	191
	55.0	35.0	2.5	7.5	0.0	40
76.	23.6	36.2	27.5	7.0	5.7	229
	21.6	34.7	30.0	7.4	6.3	190
	33.3	43.6	15.4	5.1	2.6	39
77.	37.7	35.1	18.0	5.3	3.9	228
	37.0	33.3	19.6	5.8	4.2	189
	41.0	43.6	10.3	2.6	2.6	39
78.	52.2	35.7	8.7	2.6	0.9	230
	50.3	35.5	9.9	2.1	1.0	191
	61.5	30.8	2.6	5.1	0.0	39
79.	37.8	40.9	12.6	5.2	3.5	230
	37.2	41.4	13.6	3.7	4.2	191
	41.0	38.5	7.7	12.8	0.0	39
80.	39.7	32.3	17.5	6.6	3.9	229
	38.4	33.2	19.5	4.2	4.7	190
	46.2	28.2	7.7	17.9	0.0	39
81.	43.3	38.5	13.4	2.2	2.6	231
	41.9	39.8	14.7	1.0	2.6	191
	50.0	32.5	7.5	7.5	2.5	40

	A	B	C	D	E	tot.
82.	21.2	48.9	21.6	4.8	3.5	231
	20.4	49.7	21.5	4.7	3.7	191
	25.0	45.0	22.5	5.0	2.5	40
83.	13.9	32.5	35.1	12.1	6.5	231
	12.6	32.5	35.6	12.6	6.8	191
	20.0	32.5	32.5	10.0	5.0	40
84.	46.8	45.0	6.5	0.4	1.3	231
	47.1	44.0	7.3	1.6	0.0	191
	45.0	50.0	2.5	2.5	0.0	40
85.	39.8	45.5	12.6	1.3	0.9	231
	39.8	45.5	13.1	0.5	1.0	191
	40.0	45.0	10.0	5.0	0.0	40
86.	62.2	33.0	3.9	0.9	0.0	230
	63.2	32.6	4.2	0.0	0.0	190
	57.5	35.0	2.5	5.0	0.0	40
87.	60.9	33.5	4.3	1.3	0.0	230
	59.5	35.8	4.2	0.5	0.0	190
	67.5	22.5	5.0	5.0	0.0	40
88.	33.8	43.7	15.2	4.3	3.0	231
	32.5	44.0	16.8	3.7	3.1	191
	40.0	42.5	7.5	7.5	2.5	40
89.	43.7	39.0	12.6	3.9	0.9	231
	43.5	38.7	13.6	3.7	0.5	191
	45.0	40.0	7.5	5.0	2.5	40
90.	59.3	31.2	7.8	1.3	0.4	231
	58.1	31.9	8.4	1.0	0.5	191
	65.0	27.5	5.0	2.5	0.0	40
91.	29.9	45.9	19.9	1.7	2.6	231
	28.8	48.2	18.8	1.0	3.1	191
	35.0	35.0	25.0	5.0	0.0	40
92.	11.7	32.0	39.0	9.5	7.8	231
	10.5	31.4	39.3	9.4	9.4	191
	17.5	35.0	37.5	10.0	0.0	40
93.	39.5	40.4	14.9	2.2	3.1	228
	39.2	40.2	15.9	1.6	3.2	189
	41.0	41.0	10.3	5.1	2.6	39

	A	B	C	D	E	tot.
94.	35.2	48.3	10.0	3.0	3.5	230
	33.2	50.0	11.1	2.6	3.2	190
	45.0	40.0	5.0	5.0	5.0	40
95.	49.4	33.8	11.7	3.5	1.7	231
	48.7	33.5	12.6	3.1	2.1	191
	52.5	35.0	7.5	5.0	0.0	40
96.	29.9	35.9	23.4	5.6	5.2	231
	28.3	34.0	36.2	5.8	5.8	191
	37.5	45.0	10.0	5.0	2.5	40
97.	19.5	37.7	26.8	6.5	9.5	231
	17.8	35.6	29.3	6.8	10.5	191
	27.5	47.5	15.0	5.0	5.0	40
98.	34.2	35.9	19.5	4.8	5.6	231
	31.4	36.1	21.5	4.2	6.8	191
	47.5	35.0	10.0	7.5	0.0	40
99.	34.2	39.0	17.7	4.8	4.3	231
	32.5	40.3	18.3	4.2	4.7	191
	42.5	32.5	15.0	7.5	2.5	40
100.	9.1	31.2	36.4	13.4	10.0	231
	8.9	29.3	36.1	14.1	11.5	191
	10.0	40.0	37.5	10.0	2.5	40
101.	24.6	42.5	20.6	9.6	2.6	228
	24.2	42.1	21.1	10.0	2.6	190
	26.3	44.7	18.4	7.9	2.6	38
102.	25.0	41.2	24.1	5.7	3.4	228
	22.1	40.0	27.4	5.8	4.7	190
	39.5	47.4	7.9	5.3	0.0	38
103.	44.7	44.7	8.3	1.8	0.4	228
	44.7	44.2	9.5	1.1	0.5	190
	44.7	47.4	2.6	5.3	0.0	38
104.	42.7	45.8	7.5	2.6	1.3	227
	41.8	45.5	8.5	2.6	1.6	189
	47.4	47.4	2.6	2.6	0.0	38
105.	39.5	38.5	16.2	3.5	2.2	228
	40.0	37.4	17.9	2.6	2.1	190
	36.8	44.7	7.9	7.9	2.6	38

	A	B	C	D	E	tot.
106.	32.5	35.5	19.3	6.6	6.1	228
	33.7	34.2	19.5	5.8	6.8	190
	26.3	42.1	18.4	10.5	2.6	38
107.	19.3	35.1	25.9	9.2	10.5	228
	17.4	33.2	26.8	10.0	12.6	190
	28.9	44.7	21.1	5.3	0.0	38
108.	12.3	33.3	30.7	12.7	11.0	228
	11.1	31.1	31.6	13.7	12.6	190
	18.4	44.7	26.3	7.9	2.6	38
109.	32.7	45.6	14.6	3.1	4.0	226
	31.9	44.7	16.0	3.2	4.3	188
	36.8	50.0	7.9	2.6	2.6	38
110.	27.2	44.3	16.7	4.4	7.5	228
	26.3	41.6	18.9	4.2	8.9	190
	31.6	57.9	5.3	5.3	0.0	38
111.	57.3	30.4	9.7	1.8	0.9	227
	57.1	30.2	10.6	1.6	0.5	189
	57.9	31.6	5.3	2.5	2.5	38
112.	44.1	41.0	12.3	1.3	1.3	227
	43.4	41.3	13.2	1.1	1.1	189
	47.4	39.5	7.9	2.6	2.6	38
113.	19.4	34.4	31.3	7.9	7.0	227
	18.0	35.4	31.7	6.9	7.9	189
	26.3	28.9	28.9	13.2	2.6	38
114.	42.3	45.4	10.6	1.3	0.4	227
	41.3	45.5	12.2	0.5	0.5	189
	47.4	44.7	2.6	5.3	0.0	38
115.	48.0	39.2	10.1	0.4	2.2	227
	46.0	40.2	11.1	0.0	2.4	189
	57.9	34.2	5.3	2.6	0.0	38
116.	16.3	37.0	30.8	8.4	7.5	227
	15.3	37.6	30.2	8.5	8.5	189
	21.1	34.2	34.2	7.9	2.6	38
117.	22.6	46.0	21.7	6.2	3.5	226
	23.4	43.1	23.4	5.9	4.3	188
	18.4	60.5	13.2	7.9	0.0	38

	A	B	C	D	E	tot.
118.	31.3	42.3	21.6	2.6	2.2	227
	32.3	41.8	21.7	2.1	2.1	189
	26.3	44.7	21.1	5.3	2.6	38
119.	48.9	44.5	5.3	0.9	0.4	227
	48.1	45.0	5.8	0.5	0.5	189
	52.6	42.1	2.6	2.6	0.0	38
120.	24.8	37.6	24.3	4.0	9.3	226
	23.9	36.7	26.1	3.2	10.1	188
	28.9	42.1	15.8	7.9	5.3	38
121.	41.9	41.4	12.3	3.1	1.3	227
	42.3	41.3	12.2	2.6	1.6	189
	39.5	42.1	13.2	5.3	0.0	38
122.	46.3	45.4	6.6	1.8	0.0	227
	47.6	46.0	5.8	0.5	0.0	189
	39.5	42.1	10.5	7.9	0.0	38
123.	54.0	37.2	7.5	1.3	0.0	226
	52.1	38.3	8.5	1.1	0.0	188
	63.2	31.6	2.6	2.6	0.0	38
124.	32.6	47.6	15.9	3.5	0.4	227
	30.7	48.7	16.9	3.2	0.5	189
	42.1	42.1	10.5	5.3	0.0	38
125.	35.4	29.6	19.5	11.9	3.5	226
	34.6	30.9	19.1	11.2	4.3	188
	39.5	23.7	21.1	15.8	0.0	38
126.	30.1	37.2	20.4	6.6	5.8	226
	27.1	37.8	21.8	6.9	6.4	188
	44.7	34.2	13.2	5.3	2.6	38
127.	36.1	40.1	16.7	3.5	3.5	227
	33.9	40.2	18.5	3.2	4.2	189
	42.4	39.5	7.9	5.3	0.0	38
128.	30.8	38.8	20.3	5.3	4.8	227
	29.1	40.2	20.6	4.2	5.8	189
	39.5	31.6	18.4	10.5	0.0	38
129.	43.6	42.3	7.5	4.0	2.6	227
	43.4	42.3	7.9	3.2	3.2	189
	44.7	42.1	5.3	7.9	0.0	38

	A	B	C	D	E	tot.
130.	40.5	41.4	11.0	4.0	3.1	227
	39.2	42.9	11.1	3.7	3.2	189
	47.4	34.2	10.5	5.3	2.6	38
131.	44.9	44.1	5.3	3.1	2.6	227
	45.0	44.4	4.8	2.6	3.2	189
	44.7	42.1	7.9	5.3	0.0	38
132.	20.7	41.0	26.4	6.6	5.3	227
	20.6	41.8	25.4	6.4	5.8	189
	21.1	36.8	31.6	7.9	2.6	38
133.	27.8	41.0	21.6	6.6	3.1	227
	28.6	39.7	21.7	6.3	3.7	189
	23.7	47.4	21.1	7.9	0.0	38
134.	54.2	38.3	5.7	1.8	0.0	227
	54.0	39.2	5.8	1.1	0.0	189
	55.3	34.2	5.3	5.3	0.0	38
135.	48.9	41.9	6.6	2.6	0.0	227
	47.1	43.4	6.9	2.6	0.0	189
	57.9	34.2	5.3	2.6	0.0	38
136.	30.4	50.2	11.9	5.7	1.8	227
	30.2	50.8	12.2	4.8	2.1	189
	31.6	47.4	10.5	10.5	0.0	38
137.	55.1	35.2	6.6	1.8	1.3	227
	54.5	35.4	6.9	1.6	1.6	189
	57.9	34.2	5.3	2.6	0.0	38
138.	58.6	26.4	10.6	3.1	1.3	227
	59.3	25.4	10.6	3.2	1.6	189
	55.3	31.6	10.5	2.6	0.0	38
139.	40.5	38.3	15.4	2.2	3.5	227
	40.7	37.6	15.9	1.6	4.2	189
	39.5	42.1	13.2	5.3	0.0	38
140.	35.2	35.2	18.5	6.2	4.8	227
	34.9	36.0	17.5	5.8	5.8	189
	36.8	31.6	23.7	7.9	0.0	38
141.	36.6	39.3	14.7	6.7	2.7	224
	35.5	41.4	13.4	6.5	3.2	186
	42.1	28.9	21.1	7.9	0.0	38

	A	B	C	D	E	tot.
142.	4.9	16.0	36.4	34.2	8.4	225
	4.8	16.0	36.4	33.2	9.6	187
	5.3	15.8	36.8	39.5	2.6	38
143.	2.7	5.3	29.8	50.7	11.6	225
	2.7	3.7	32.1	49.2	12.3	187
	2.6	13.2	18.4	57.9	7.9	38
144.	13.3	41.3	23.6	16.4	5.3	225
	12.8	40.6	24.6	15.5	6.4	187
	15.8	44.7	18.4	21.1	0.0	38
145.	11.6	34.2	27.6	20.0	6.7	225
	10.2	32.1	30.5	19.3	8.0	187
	18.4	44.7	13.2	23.7	0.0	38
146.	8.9	22.8	38.4	23.2	6.7	224
	9.6	20.3	40.1	22.5	7.5	187
	5.4	35.1	29.7	27.0	2.7	37
147.	4.0	14.7	32.9	36.4	12.0	225
	3.7	14.4	33.7	34.8	13.4	187
	5.3	15.8	28.9	44.7	5.3	38
148.	4.5	13.8	35.7	34.4	11.6	224
	4.3	12.8	37.4	32.6	12.8	187
	5.4	18.9	27.0	43.2	5.4	37
149.	6.3	19.6	34.4	29.5	10.3	224
	7.0	17.6	35.8	27.8	11.8	187
	2.7	29.7	27.0	37.8	2.7	37
150.	7.1	19.6	34.4	28.1	10.7	224
	7.5	20.3	34.8	26.2	11.2	187
	5.4	16.2	32.4	37.8	8.1	37
151.	1.3	3.6	27.6	52.9	14.7	225
	1.1	2.7	28.3	51.9	16.0	187
	2.6	7.9	23.7	57.9	7.9	38
152.	3.1	17.8	31.6	37.8	9.8	225
	3.7	16.0	32.1	37.4	10.7	187
	26.3	28.9	39.5	5.3	0.0	38
153.	9.3	38.7	29.8	18.2	4.0	225
	10.2	38.0	29.9	17.1	4.8	187
	5.3	42.1	28.9	23.7	0.0	38

	A	B	C	D	E	tot.
154.	17.9	40.2	26.3	12.1	3.6	224
	16.1	39.8	28.5	11.8	3.8	186
	26.3	42.1	15.8	13.2	2.6	38
155.	3.7	16.6	28.1	39.2	12.4	217
	2.2	15.6	29.4	37.8	15.0	180
	10.8	21.6	21.6	45.9	0.0	37

APPENDIX D. RANK ORDER OF EVALUATIVE CRITERIA BY FREQUENCY AND WEIGHTED FREQUENCY

Preference ratings of evaluative criteria

Rank	By Frequency	By Weighted Frequency
Personal Characteristics: Criterion numbers (score)		
1.	5 (108)	5 (256)
2.	7 (102)	7 (202)
3.	3 (80)	3 (187)
4.	9 (33)	9 (56)
5.	11 (28)	11 (49)
6.	14 (22)	14 (35)
7.	6 (18)	6 (30)
8.	2, tie (15)	4 (29)
9.	4	2 (26)
10.	1, tie (10)	10 (18)
11.	13	1 (15)
12.	10 (9)	13 (14)
13.	16 (6)	16 (9)
14.	12 (2)	12 (2)
15.	8, tie (1)	8, tie (1)
16.	15	15
Professional Qualities: Criterion numbers (score)		
1.	25 (64)	25 (234)
2.	27 (42)	27 (93)
3.	42 (32)	42 (63)
4.	26 (27)	26 (54)
5.	33 (26)	33 (52)
6.	46 (25)	24 (46)
7.	24, tie (24)	28 (41)
8.	41	46 (38)
9.	32 (20)	32 (38)
10.	28 (19)	21 (35)
11.	21 (18)	42 (33)
12.	30 (14)	30 (27)
13.	18, tie (12)	18 (23)
14.	48	37 (21)
15.	29 (11)	29 (19)
16.	37 (10)	22, tie (16)
17.	22 (9)	48
18.	45 (7)	17, tie (11)
19.	17,	45
20.	34, tie (5)	36 (9)
21.	36	23,
22.	23,	34, tie (7)
23.	35, tie (4)	35 (6)
24.	40	40,
25.	43,	43, tie (5)
26.	44, tie (3)	44
27.	49 (3)	47 (8)
28.	47 (2)	49 (3)
29.	19,	31 (2)
30.	20,	19,
31.	31, tie (1)	20, tie (1)
32.	39	39
33.	38 (0)	38 (0)
Classroom Activities: Criterion numbers (score)		
1.	53 (47)	53 (108)
2.	57 (38)	51 (80)
3.	51 (34)	57, tie (76)
4.	62 (31)	62
5.	52 (25)	52 (55)
6.	87 (23)	87 (46)
7.	67, tie (19)	67 (33)
8.	75	65 (32)
9.	65 (18)	50, tie (28)
10.	72 (16)	75
11.	86 (15)	58,
12.	58 (13)	72, tie (26)
13.	50, tie (12)	86
14.	78	68, tie (23)
15.	63,	78
16.	68, tie (10)	77 (20)
17.	71	63, tie (18)
18.	77, tie (9)	84
19.	84	71 (17)
20.	55,	55, tie (12)
21.	73, tie (6)	90
22.	103	73 (11)
23.	90, tie (5)	66 (10)
24.	104	76,
25.	56,	103, tie (9)
26.	66, tie (4)	104,
27.	69	105
28.	60,	81 (7)

Rank	By Frequency	By Weighted Frequency
Classroom Activities (continued): Criterion numbers (score)		
29.	61	64
30.	76	99 tie (6)
31.	81	56
32.	85 tie (3)	69
33.	94	85 tie (5)
34.	99	93
35.	105	94
36.	59	60
37.	64	61 tie (4)
38.	79	79
39.	80	54
40.	82 tie (2)	59
41.	93	82 tie (3)
42.	95	101
43.	101	106
44.	106	70
45.	54	80
46.	70	89
47.	74	95 tie (2)
48.	89	96
49.	96	110
50.	97 tie (1)	74
51.	98	97
52.	107	98 tie (1)
53.	109	107
54.	110	109
55.	83	83
56.	88	88
57.	91	91
58.	92 tie (0)	92 tie (0)
59.	100	100
60.	102	102
61.	108	108

Rank	By Frequency	By Weighted Frequency
Interpersonal Relations: Criterion numbers (score)		
1.	111 (84)	111 (225)
2.	115 (59)	115 (133)
3.	119 (43)	119 (76)
4.	121 tie (27)	123 (60)
5.	123	121 (56)
6.	135 (18)	114 (33)
7.	112 (16)	112 (32)
8.	134 (15)	134 (30)
9.	114 tie (14)	122 (28)
10.	137	135 (26)
11.	122	133
12.	133 tie (13)	137 tie (23)
13.	139	138
14.	138 (12)	139 (16)
15.	129 (11)	129 (14)
16.	124 (10)	124 (13)
17.	130 (9)	130 tie (12)
18.	131 tie (8)	131
19.	140	140 (11)
20.	120	118 tie (10)
21.	132 tie (5)	132
22.	128	117
23.	117 tie (5)	128 tie (9)
24.	118	120
25.	127 tie (4)	127 tie (8)
26.	136	136 (6)
27.	126 (3)	126 (5)
28.	113 (2)	113 (3)
29.	116 tie (1)	125 (2)
30.	125	116 (1)

	By Frequency	By Weighted Frequency
Product Measures: Criterion numbers (score)		
1.	141 (119)	141 (326)
2.	154 (68)	154 (124)
3.	144 (56)	144 (102)
4.	153 (49)	153 (84)
5.	145 (37)	145 (61)
6.	152 (14)	152 (49)
7.	146 (22)	146 (42)
8.	149 (15)	149 (21)
9.	142 (10)	142 (20)
10.	148 (9)	148 (16)
11.	143 (7)	143 (9)
12.	150 (5)	150 (7)
13.	155 (3)	155 (3)
14.	147 (1)	147 (2)
15.	151 (0)	151 (0)